SYMPOSIUM

ON

ELECTROMAGNETIC WAVE THEORY

—

Report of Conference
held
June 20-25, 1955
at the
University of Michigan

—

Sponsored by
Commission VI of URSI
and the
University of Michigan

IRE TRANSACTIONS ® PGAP IS A QUARTERLY PUBLICATION DEVOTED TO EXPERIMENTAL AND THEORETICAL PAPERS ON ANTENNAS AND WIRELESS PROPAGATION OF ELECTROMAGNETIC WAVES

PROCEEDINGS OF THE SYMPOSIUM ON ELECTROMAGNETIC WAVE THEORY

THE UNIVERSITY OF MICHIGAN

Ann Arbor, Michigan

June 20 – June 25, 1955

Organizing Committee

K. M. Siegel, *Chairman*
Head, Theory and Analysis Department, Engineering Research Inst., The University of Michigan.

W. G. Dow
Professor, Electrical Engineering, The University of Michigan

G. Hok
Professor, Electrical Engineering, The University of Michigan

J. D. Shortt, Jr.
Assistant to the Director of University Relations, The University of Michigan

S. S. Attwood
Professor, Electrical Engineering, Chairman, Electrical Engineering Dept., The University of Michigan

R. V. Churchill
Professor, Mathematics, The University of Michigan

D. M. Dennison
Chairman, Dept. of Physics, The University of Michigan

C. L. Dolph
Associate Professor, Mathematics, The University of Michigan

S. Silver
Professor of Engineering Science, University of California, Berkeley, California (*ex-officio* member for Commission VI of URSI)

N. Smith
Project Supervisor, Engineering Research Inst., The University of Michigan

G. E. Uhlenbeck
Professor, Physics, The University of Michigan

H. W. Welch, Jr.
Associate Professor, Electrical Engineering, The University of Michigan

Canadian Sub-Committee

G. Sinclair, *Chairman*
University of Toronto

G. Bekefi
McGill University

G. A. Miller
National Research Council

J. C. W. Scott
Defense Research Board

Arrangements Committee

K. M. Siegel, *Chairman*
Head, Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

J. W. Crispin, Jr., *Secretary*
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

W. E. Burdick
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

C. F. Carlson
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

J. D. Shortt, Jr.
Asst. to the Director of University Relations, The University of Michigan

Proceedings Committee

K. M. Siegel
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

T. B. Curtz, *Chairman*
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

K. A. Brooks
Theory and Analysis Dept., Engineering Research Inst., The University of Michigan

Experimental Demonstration Committee

B. Levy, *Co-Chairman*
The University of Michigan

J. M. Wolf, *Co-Chairman*
The University of Michigan

TABLE OF CONTENTS

Introduction

KEEVE M. SIEGEL

FROM June 20 through June 25, 1955, an International Symposium on Electromagnetic Wave Theory was held at Ann Arbor, Michigan. It was sponsored by Commission VI of URSI (The International Scientific Radio Union) and the Chrysler Corporation, Convair, Ford Motor Company, General Motors Corporation, Hughes Aircraft Company, Melpar, Inc., Philco Corporation, Republic Aviation Corporation, Republic Steel Corporation, Sylvania Electric Products, Inc., United States Air Force, United States Army, United States Navy, Westinghouse Electric Corporation, and The University of Michigan.

The purpose of this symposium was exclusively educational. The aim of the University of Michigan Organizing Committee was to assemble the leading scientists in the field of electromagnetic wave theory, in order to provide the opportunity for them to receive and present the latest developments. It was our belief that the discussions and stimuli furnished by such a meeting could shorten by many years the time required for major scientific advances in boundary value problems of diffraction and scattering theory, forward and multiple scattering, antenna theory and microwave optics, and propagation in doubly refracting media.

This volume contains the invited papers and abstracts of the contributed papers that were presented at the symposium. In addition, the panel discussions which followed each of the sessions have been summarized by the panel chairmen and are included in the text of the proceedings which follows.

I am taking this opportunity to thank the Professional Group on Antennas and Propagation of the Institute of Radio Engineers for publishing the proceedings of this symposium. I would also like to express my gratitude to the large number of scientists without whose help and encouragement this Symposium would not have been possible. It is my sincere hope that the members of the Professional Group will find these proceedings as stimulating and interesting as I have.

Welcoming Address

SAMUEL SILVER

MR. CHAIRMAN, President Hatcher, Ladies and Gentlemen: Commission VI of the International Scientific Radio Union enjoys here a dual position of being both a guest and a host. The concept and plans for the Symposium originated in the University of Michigan, and the realization of those plans is due almost wholly to the efforts of the Organizing Committee of the University. When the invitation was extended to the International Commission on Radio Waves and Circuits to participate as scientific sponsors, we accepted readily and our view was supported enthusiastically by the President of the Union, Father Lejay, and the executive board. To us this invitation presented an opportunity to implement a major phase of our program in applied electromagnetic theory. I wish at this time, President Hatcher, to express to you and the Organizing Committee our appreciation for calling on us to join in and partake of this meeting.

The purposes of U.R.S.I. are to promote international exchange of information and collaboration in research, and to highlight important developments and currently outstanding problems in the various fundamental fields underlying the technical aspects of radio and communications. It is clear that these purposes are to be served well this week. It is my privilege to express on behalf of Commission VI its welcome to this distinguished gathering. I also must express my personal pleasure in seeing so many of my colleagues again and in looking forward to making so many more.

As one views the extensive scope of the program, he cannot help but be impressed by the renaissance which has taken place in the field of classical electromagnetic theory. Some fifteen years ago a symposium of this magnitude in this field could hardly be visualized. Electromagnetic theory in the classical domain seemed to have run its course and in many quarters the subject resided in closed books and in closed minds. Today we witness evidence of a vital and engaging field. "Solutions in principle" have shown themselves to be inadequate and indeed in a number of instances the principle became fuzzy on more critical examination.

Diffraction and scattering theory has grown almost into a field of its own, laying new ground for a diversity of special researches in microwave optics, antennas, and propagation. The search for a deeper understanding of the physical phenomena has led to an amplification of fundamental ideas, and has revealed the weakness of systems of concepts which we have long considered to be well known and obvious. I need only mention in this connection geometrical optics and the general theory of asymptotic solutions of the field equations.

Much progress has been made recently and more is to be reported at this meeting; much yet remains to be done. The questions are not merely pertinent to technical applications but are of a fundamental character concerning the nature of the electromagnetic field. The problems that concern us here are significant for all phenomena to which the wave equation is to be applied in analysis, and the results of the study of the classical field problem may yet have ramifications in the atomic and nuclear domain.

The mathematician as well as the physicist has fertile ground of fundamental research to till in our area. Diffraction theory has presented us with new types of integral equations and a variety of generalized variational principles. There is progress yet to be made in fusing these into a coherent system. Propagation in isotropic homogeneous media poses a multitude of difficult problems. Inasmuch as mathematical solutions are incomplete, the interpretation of the physical phenomena remains open to question. The situation is made even more complicated when dealing with a medium such as the earth's atsmosphere which, following the normal law of the perversity of nature, assumes in relatively short periods the whole variety of simple models which the scientist conjures up to be the subject of his analysis.

The ferrites, the outgrowth of a renaissance in yet another field—solid state physics—were soon recognized to possess great potentialities for technical developments. We quickly came to realize how little we actually know about propagation in anisotropic homogeneous media. More is required than just a discussion of a homogeneous plane wave in the infinite medium. There is an urgent need for extensive solutions to interior boundary value problems, such as deal with propagation in waveguides containing ferrites, in order to make possible the fullest utilization of the technical potentialities. The exterior problems are also assuming importance, though the medium in this case may be a magneto-ionic one rather than a ferrite, but the fundamental nature of the problem is not radically affected thereby.

There is no need to develop the theme that the renewed vitality of classical electromagnetic theory stems from the technical developments of the past fifteen years. The point is to be made, however, in relation to a most significant aspect of this symposium; namely, that engineers, physicists and mathematicians are meeting here together for an intimate exchange of views. The excellent program provides grist for the mill and, out of this confluence of interests and ideas, there cannot help but come fruitful results in solutions to existing problems as well as the creation of new ones.

On Field Representations in Terms of Leaky Modes or Eigenmodes

N. MARCUVITZ[†]

Summary—Solutions to source-excited field problems are frequently represented as superpositions of source-free field solutions. The latter are in general of two types: eigenmodes and noneigenmodes which are related to the zeros of the total impedance or alternatively the poles of the scattering coefficient of a system. The eigenmodes are everywhere finite and comprise a complete orthogonal set. The noneigenmodes become infinite in the infitely remote spatial limits of a region and are not in general members of a complete orthogonal set; examples are "radio-active states," "damped resonances," and "leaky waves." Despite their physically singular behavior, the nonmodal solutions can be employed to represent field solutions in certain ranges.

A FAMILIAR procedure in mathematical physics for solving field equations in the presence of sources involves the determination of the possible source-free field solutions and their subsequent superposition. In regions of finite extent bounded by impermeable walls; i.e., "closed regions," the source-free solutions generally possess orthogonality and completeness properties that permit an arbitrary function, and in particular a desired field solution, to be represented by their superposition. Such source-free solutions are termed the characteristic (eigen) or normal modes of the given region; they possess a discrete spectrum, are everywhere finite, and individually satisfy the field equations plus appropriate boundary conditions. In regions of infinite extent; i.e., "open regions," there may exist a corresponding discrete spectrum of modes but for completeness these must be supplemented in general by a continuous spectrum of characteristic modes to permit the representation of an arbitrary function. In contrast to the discrete modes, the continuous modes are improper in that *individually* they are not absolute square integrable (with finite energy) nor do they satisfy the requisite boundary conditions at the singular point, infinity. In the presence of a continuous spectrum there may also exist a set of nonmodal solutions of the source-free field equations which in distinction to the modes become infinite in the infinitely remote limits of an open region; in this category belong the so-called "radio-active state," "damped resonance," and "leaky wave" solutions. Despite their physically unacceptable behavior in the given region, such nonmodal solutions may nevertheless be employed for the representation of field solutions in an extended region of which the given region is a subspace.

In the following we shall attempt to delimit the usefulness of the nonmodal solutions, and their relation to the modal solutions as well. Although not in general members of a complete set of orthgonal functions, the nonmodal solutions may nevertheless be employed to obtain rapidly convergent field representations in certain ranges. A pragmatic characterization of both the nonmodal and the discrete modal solutions stems from their connection with the zeros of an impedance of a system. As is well known, the zeros of the total impedance[1] (admittance) or, alternatively, the poles of the scattering coefficient of a system represent "resonant frequencies" that distinguish all possible source-free solutions. For a conservative (hermitian) system real resonant frequencies represent eigenmodes whereas complex resonant frequencies characterize the nonmodal solutions. In the case of nonconservative; i.e., dissipative, systems both the modal and nonmodal frequencies are in general complex and one of the points to be discussed is how they are to be distinguished.

Two types of examples will be treated. In the first we consider a spatial volume containing open cavities in the electromagnetic and acoustic case, or regions of nonvanishing potential in the quantum-mechanical case. Such regions will be regarded as three dimensional subspaces of a four-dimensional space-time. In a nondissipative system the invariance of the field equations of interest under a time displacement implies the existence of source-free solutions with a time dependence exp $(-i\omega t)$. The determination of the possible values of ω (or E/t) and the associated spatial field distributions pose a field problem the solutions of which may be of the modal or nonmodal type, admitting, respectively, real or complex values of ω. The latter correspond to damped oscillations that become infinite at the infinitely distant spatial limits. In the acoustic and the electromagnetic field these nonmodal oscillations are designated as "damped resonances," in the quantum-mechanical field as "radioactive states." Typical nonmodal solutions of this class have the asymptotic form

$$e^{i w_i r/e} e^{-i w_i t} \tag{1}$$

with Im $\omega_i < o$.

A second type of example will be concerned with steady-state acoustic, electromagnetic, etc., field problems associated with the two-dimensional cross sections of open cylindrical waveguides. Such a waveguide with

---

† Polytechnic Institute of Brooklyn, Brooklyn, N.Y.

[1] A quantity proportional to the logarithmic space derivative of a wave function.

axis along z possesses translational invariance along the axial direction, whence there is implied the existence of source-free field solutions with z dependence $\exp(i\beta z)$. The determination of the permitted values of β and of the corresponding cross-sectional field distributions poses a cross-sectional field problem whose solutions may be either of the modal or nonmodal type. For a nondissipative medium the modal solutions correspond to real or imaginary values of β (i.e., real values of β^2) and are representative of waves either propagating with undiminished amplitude or evanescent with invariable phase variation along the z-axis. On the other hand, the nonmodal solutions correspond to complex values of β and are indicative of "leaky waves" either propagating with diminishing amplitude or evanescent with variable phase variation along the z direction; as noted above, the amplitudes of these waves increase indefinitely in one of the infinitely remote reaches of the guide cross section. Typical of a nonmodal wave would be a solution of the asymptotic form

$$e^{i\sqrt{k^2 - \beta_i^2}\,y}\, e^{i\beta_i z} = e^{i\,k\,r\,\cos\,(\Phi - \Phi_i)}, \quad (2)$$

with $\mathrm{Im}\sqrt{k^2 - \beta_i^2} < 0$ and $\mathrm{Im}\,\beta_i > 0$; the cross-sectional coordinate is $y = r\cos\Phi$ and $\beta_i = k\sin\Phi_i$.

In the above examples the subspace of interest was extended by displacement along either a temporal or spatial direction in such a manner as to insure the existence of field solutions with exponential dependence of "periodicity" ω or β along the displacement direction. If, consistent with this field periodicity, a point or line source with periodicity ω or β is inserted into these three or two dimensional regions, there will be excited a field whose amplitude will be ω or β dependent. This field, represented by the so-called characteristic Green's function or resolvent of the given region, will become singular at those values of ω or β for which there can exist source-free solutions, i.e., resonances. These resonances correspond to either modal or nonmodal solutions. The characteristic Green's function provides a well-defined means for distinguishing between those singular values of β or ω that correspond to the modes of a complete orthogonal set and those that do not. The singularities of this Green's function take the form of either poles or branch cuts. The latter are generally present only in the case of open structures, in which event it becomes necessary to distinguish between proper and improper β or ω branches of the characteristic Green's function. The nonmodal singularities lie in the improper branch.

There exists an established formalism for the construction of Green's functions and for the determination of a complete spectrum of modes from the characteristic Green's function. Since this formalism will be employed in the examples treated below, we shall present a summary at this point. In many regions the determination of the

source-free fields is reducible to the solution of a one-dimensional second-order differential equation. Accordingly, we consider first that homogeneous equation

$$\left[-\frac{d}{dx}\,p(x)\,\frac{d}{dx} + q(x) - \lambda_a\,w(x) \right] \Psi_a(x) = 0, \quad (3)$$

where p, q, w are piecewise continuous functions and λ_a and $\psi_a(\mathrm{x})$ are to be determined subject to certain boundary conditions stated below. There is a corresponding inhomogeneous problem for finding the field produced by a point source at $x = x'$; *for arbitrary complex* λ, this serves to define the characteristic Green's function $g(x,x';\lambda)$ as

$$\left[\frac{d}{dx}\,p(x)\,\frac{d}{dx} + q(x) - \lambda w(x) \right] g(x,x';\lambda) = \delta(x - x'), \quad (4)$$

where $\delta(x-x')$ is the usual delta function and, as in (3), the required boundary conditions will remain temporarily unspecified. As mentioned above, $g(x,x';\lambda)$ possesses pole singularities at those values $\lambda = \lambda_a$ for which there exist source-free solutions $\psi_a(x)$ satisfying the required boundary conditions. In open (infinite x-regions), $g(x,x';\lambda)$ generally also possesses branch point singularities; one is led thereby to choose branch cuts in the λ-plane so as to obtain a "proper" branch on which $g(x,x';\lambda) \to 0$ as $\lambda \to \infty$. The connection between the proper mode solutions $\psi_a(x)$ of (3) and $g(x,x';\lambda)$ is then provided by the completeness relation

$$\delta(x - x') = -\frac{1}{2\pi i} \oint g(x, x'; \lambda)\, d\lambda$$

$$= \sum_a \Psi_a(x)\, \Psi_a(x'), \quad (5)$$

where the contour integral encloses *all* singularities of $g(x,x';\lambda)$ *in the proper branch* of the λ-plane The sum on the right is intended to represent both the residue contributions from the poles λ_a and the branch cut contributions as well; hence the sum sign denotes both a summation over discrete spectral points λ_a and an integral over the continuous spectral values of λ corresponding to the branch cuts.

The explicit evaluation of the Green's function $g(x,x') = g(x,x';\lambda)$ presupposes the ability to obtain, for arbitrary λ, two independent solutions of the homogeneous form of (4). Let $c(x,x_0)$ and $s(x,x_0)$ be two such solutions distinguished at $x = x_0$ by the conditions

$$\begin{array}{cc} \mathrm{c}(\mathrm{x_0, x_0}) = 1 & \mathrm{s}(\mathrm{x_0, x_0}) = 0 \\ \mathrm{p}(\mathrm{x_0})\,\mathrm{c'}(\mathrm{x_0, x_0}) = 0 & \mathrm{p}(\mathrm{x_0})\,\mathrm{s'}(\mathrm{x_0, x_0}) = 1, \end{array} \quad (6)$$

where the prime denotes differentiation with respect to x. In consequence of the λ-independent conditions (6), $c(x,x_0)$ and $s(x,x_0)$ are integral functions of λ and nonsingular in the finite part of the λ-plane. The presence of a point source at x' gives rise to two waves, $u^+(x)$ for $x > x'$ and $u^-(x)$ for $x < x'$, the former satisfying the

prescribed boundary conditions at the lower limit $x = a$ and the latter at the upper limit $x = b$. To within an amplitude factor these homogeneous wave solutions may be represented as

$$u^+(x) = c(x,x_0) + X^+(\lambda)\, s(x,x_0)$$
$$u^-(x) = c(x,x_0) + X^-(\lambda)\, s(x,x_0), \qquad (7)$$

where in view of (6) the logarithmic derivatives $X^+ = p(x_0)\, u^{+\prime}(x_e)/u^+(x_0)$ and $X^- = p(x_0)\, u^{-\prime}(x_0)/u^-(x_0)$ (which to within a factor of $\sqrt{-1}$ play the roles of impedances (or admittances at x_0 looking, respectively, in the directions of increasing and decreasing x) are determined by the required boundary conditions. The characteristic Green's function may then be represented discontinuously by

$$g(x, x') = -\,\frac{u^-(x_<)\, u^+(x_>)}{X^-(\lambda) + X^+(\lambda)}, \qquad (8)$$

where $x <$ stands for the smaller and x or x', and $x >$ the larger.

It is the twofold purpose of this report to emphasize through a number of examples:

1. The distinction between the modal and nonmodal source-free solutions of various field problems.

2. The existence in open systems not only of spectral representations of source excited field in terms of eigenmodes, but also of nonspectral and frequently more rapidly convergent representations in terms of the singular nonmodal solutions.

The Interpretation of Numerical Results Obtained by Rigorous Diffraction Theory for Cylinders and Spheres [*]

H. C. VAN DE HULST[†]

Summary—The classical solutions of the scattering problems for homogeneous spheres and circular cylinders are taken as the starting point for a physical discussion. The limiting cases arising if two or three of the parameters x, m, and $x(m$-$1)$ are very small or very large are surveyed and interpreted. The remaining paper deals with bodies fairly large compared to the wavelength. It is shown that exact transformations and/or approximate theories may help in the problem of interpolating between rigorous results. It is also shown that the extinction by large bodies is due to a combination of the classical effects of diffraction and geometrical optics with the less familiar edge effects and surface waves. The sign and magnitude of the edge effects for bodies of different refractive index admits of a simple explanation.

INTRODUCTION

WHY DISCUSS again a problem that was rigorously and completely solved fifty years ago? The answer is: to satisfy the needs of the practical physicist, engineer, chemist, or astronomer. The diffraction of a plane electromagnetic wave by a homogeneous sphere was solved by Mie in 1908. The diffraction of a plane electromagnetic wave by a homogeneous circular cylinder of infinite length with its axis perpendicular to the direction of propagation of the incident wave was solved by Rayleigh in 1881. Since then a multitude of numerical results have been computed from these formulas, partly for mathematical curiosity, partly for practical use. The parameters entering into these solutions are $x = ka = 2\pi a/\lambda$ (where a = radius, k = wave number in outer medium, λ = wavelength in outer medium) and m = relative refractive index of inner medium with respect to outer medium. The distant field (scattered radiation) depends on one angle θ, which will be measured from the direction of propagation of the incident wave. Numerical results are now available for spheres with x up to 400 and many values of m, real and complex. Results for cylinders are available; e.g., for $m = \infty$, 1.5, 1.25 and $\sqrt{2}(1 - i)$ and upper limits of x ranging from 4 to 10.

These results, although very useful, do not satisfy the practical needs. The following requests go beyond what is given by the present tables:

1. Results for other geometrical forms; e.g. ellipsoids;
2. Results for intermediate values of x, m, and θ;
3. Results for a mixture of sizes, for white light, or integrated over a range of angles;
4. Results for bodies that are only roughly spherical or roughly homogeneous.

Point (1) will be amply discussed at this symposium and may be passed here. Point (2) is important mainly as a preliminary to (3). The importance of point (3) is not obvious. It might seem that this is a practical detail to be left to a computer. Yet this is not advisable for large sizes. The rigorous results show all minute interference effects between rays emerging in the same direction on different optical paths through the body. The interference fringes shift rapidly with slight variations of x and m and instead of integrating these maxima and minima it may be simpler in most cases to forget about interference and just add intensities. For this reason the results meant sub (3) lie on much smoother curves than the familiar rigorous results. And for precisely that reason they are more useful. Point (4), finally, is obvious. A rain-drop is a good approximation to a smooth sphere but an interstellar dust grain is not thought of as a perfect sphere or ellipsoid. Nevertheless we need good estimates of its light scattering.

In order to meet these requests, a great deal can be gained by a careful discussion of the limiting cases and asymptotic formulas. All of these lead to expressions that make it more practical to cover the range of parameters completely. On the basis of such a discussion it may also be possible to develop an intuitive understanding of the problem that makes it possible to answer some questions posed sub (4). For instance: How is the rainbow affected if the drops are slightly elliptical and how if the drop surface is corrugated or rippled? The exact answer would be difficult to find but anybody who knows Descartes' theory of the rainbow can venture to answer that the first effect would slightly displace the rainbow and that the second effect would make it vanish almost completely.

A SURVEY OF LIMITING CASES

The combinations of the parameters m and x may be displayed as points in a plane. Simple limiting cases occur when two of the numbers

$$x, \; m, \; \text{and} \; x(m - 1)$$

have absolute values that are either very small or very large compared to 1. They form the sides of the hexagon in Fig. 1. Even simpler formulas result in the overlapping regions where all three numbers satisfy such an inequality.

As an example, Table 1 shows the expressions of the efficiency factor for extinction, Q, for spheres and for very long circular cylinders on perpendicular incidence
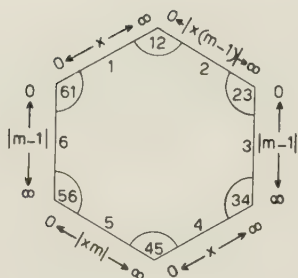
Fig. 1—A survey of limiting cases.

of linearly-polarized radiation. The factor Q is defined as the ratio of the extinction cross section to the geometrical cross section. In this table m is supposed real and >1 in order to exclude absorption losses.

TABLE I
Expressions of the Efficiency Factor Q

Corner of Fig. 1	Spheres	Cylinders Case I: E // axis	Cylinders Case II: H // axis
61	$\frac{32}{27}(m-1)^2 x^4$	$\frac{\pi^2}{2}(m-1)^2 x^3$	$\frac{\pi^2}{4}(m-1)^2 x^3$
12	$2(m-1)^2 x^2$	$\frac{8}{3}(m-1)^2 x^2$	$\frac{8}{3}(m-1)^2 x^2$
23	2	2	2
34	2	2	2
45	$\frac{10}{3}x^4$	$\frac{2}{x}\left\{1+\left(\frac{2s}{\pi}\right)^2\right\}^{-1}$	$\frac{3\pi^2}{8}x^3$
56	$\frac{8}{3}x^4$	$\frac{\pi^2}{8}m^4 x^3$	$\frac{\pi^2}{4}x^3$

$s = 0.5771 + \ln(x/2)$

A similar table might be drawn up for radar cross sections, and for absorption (in the case of complex m), and so on.

In trying to obtain numerical results for a point inside the hexagon it is advisable to look for a way to approach it from the nearest side. Mathematically, this means that we have to find an asymptotic formula in which the result for that side is the main term. Physically, it means that different descriptions are adequate in different regions of the $m - x$ domain. The main effects may be described as follows.

side 1: Independent dipole scattering by the separate volume elements. The interference of the scattered waves determines a simple form factor.
side 2: The waves pass nearly straight through the body and the scattering pattern may be derived by Huygens' principle applied to the emerging wave front.
side 3: Geometrical optics.
side 4: Perfect conductors; simplified boundary conditions.
side 5: Resonance of the body's modes.
side 6: Dipole scattering.

The modification of the formulas necessary to derive numerical results for values of the parameters that do not amply satisfy the inequalities is simplest near side 6.
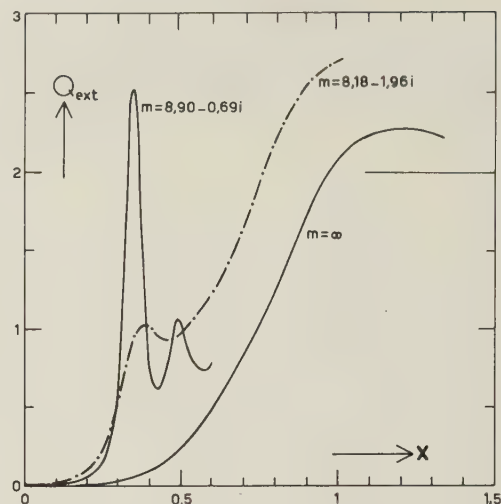
Fig. 2—Extinction curves of spheres for two complex values of the refractive index m (after Lowan) and for $m = \infty$. Co-ordinates: $Q =$ efficiency factor, $x = 2\pi a/\lambda$, where $a =$ radius and $\lambda =$ wavelength

It is the inclusion of higher multipole scattering, that gives the classical solutions. These converge everywhere but require for their evaluation few terms for small x and many for larger x. They are also quite sufficient for a discussion of the resonance effects. Fig. 2 shows the extinction factors[1] for spheres with values of

$$m = n - in'.$$

The magnetic dipole resonance is clearly seen at $x = \pi/n$. The peak is stronger for the smaller value of n', which indicates smaller damping.

INTERPOLATION PROBLEMS

In 1909 Debye[2] indicated that it is possible to show by means of asymptotical expansions that the solutions just mentioned approach in the limit of very large x the result derived from geometrical optics. He also mentioned that the neglected terms are of the order of $x^{-2/3}$ times the main term. A practical method to compute these terms was not given, so it seemed virtually impossible to enter the hexagon from the opposite side (side 3). Improving a method first used by van der Pol and Bremmer, Franz has recently found a practical way to do so but no numerical results are yet available.

As long as this is true, we have to use numerical results obtained by much labor (or by a big machine) from the classical solutions. This poses a serious interpolation problem. Both the intermediate results (the coefficients of the series) and the final results (the scattered intensities, the radar and extinction cross sections) show weird fluctuations. Unless the computed points lie sufficiently close together to detect the nature of these fluctuations empirically, we are very much in need of a theoretical notion of these fluctuations. Even a rough idea may help to avoid errors in the interpolation. This point will be made clear by two examples.

[1] A. N. Lowan, "Tables of Scattering Functions for Spherical Particles," NBS Appl. Math. Ser. 4; 1949.
[2] P. Debye, *Ann. der Phys.* vol. 30, p. 59; 1909.
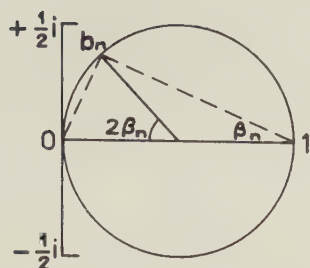
Fig. 3—Locus of complex coefficients.

1. It is obviously desirable to have a check on the coefficients of the expansions. The efficiency factor for extinction by long circular cylinders in Case I (electric field // axis) is given by

$$Q_{\text{ext}} = \frac{2}{x} \sum_{n=-\infty}^{\infty} \text{Re } b_n,$$

where

$$b_n = \frac{mJ_n'(y)\, J_n(x) - J_n(y)\, J_n'(x)}{mJ_n'(y)\, H_n(x) - J_n(y)\, H_n'(x)}.$$

Here $J_n(x)$ represents the Bessel function, $H_n(x)$ the second Hankel function, $y = mx$, primes denote derivatives, and Re means "real part of." The efficiency factor for scattering is

$$Q_{\text{sca}} = \frac{2}{x} \sum_{n=-\infty}^{\infty} |b_n|^2.$$

The fact that Q_{ext} has to equal Q_{sca} for a nonlossy material; i.e. m = real, suggests that $|b_n|^2 = \text{Re } b_n$ from which follows at once that these coefficients lie on a circle in the complex domain (Fig. 3) and that a transformation to the real angle β_n by means of

$$b_n = \tfrac{1}{2}(1 - e^{-2i\beta_n})$$

would be in order. The same formulas with complex coefficients a_n and angles α_n hold for long circular cylinders in Case II (magnetic field // axis). Other coefficients a_n' and b_n' that allow of the same trans-, formation, appear in the solution for the sphere. The original Mie coefficients, that have been used by most authors had additional factors $\pm(2n + 1)i^{2n+1}$, that make it even more difficult to see any regularity.

It is obvious that a graphical interpolation, or check, is more easily made with the real angles than with the complex coefficients. When such a plot was first made, unexpected help came from the remarkable properties of the curves in the "nodes," where $\alpha_n = \beta_n$ (Fig. 4). The curves of β_{n-1} and β_{n+1} pass through the nodes indicated by squares and defined by $J_n(y) = 0$. They have a common tangent with the curve of α_n in the node, while the curve β_n has a horizontal tangent. These properties were read from the graphs and later verified by some juggling with the Bessel functions. The positions of all nodes and the slopes of the curves in the nodes follow from simple formulas. When these have been constructed, only a few additional points have to be
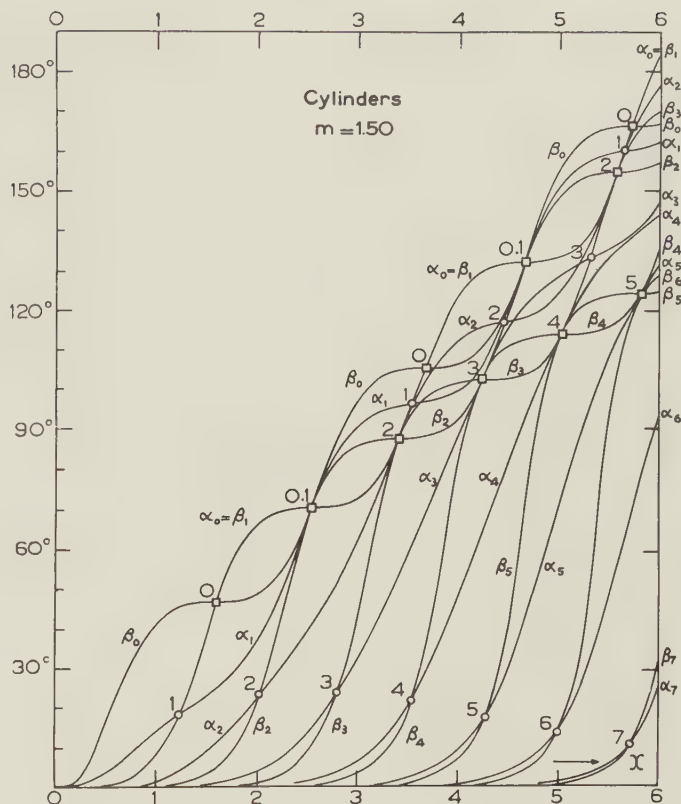
Fig. 4—Phase angles α_n and β_n for circular cylinders with $m = 1.50$. The two kinds of nodes are indicated by circles and squares.

computed to draw the full curves with reasonable accuracy. Also computation errors are readily detected.

This figure is presented here in the hope that similar transformations may be useful in more complicated problems; e.g. in the diffraction by ellipsoids. The properties would have been missed if the figure had not been drawn. If this can happen with formulas containing Bessel functions it certainly can happen with formulas containing spheroidal functions.

2. Gumprecht, Sung, Chin and Sliepcevich[3] have given the complex amplitudes of the radiation scattered by spheres with $m = 1.33$ (water drops at visual wavelengths) and values of x ranging with steps of 5 from 10 to 40. In a different context, an integration over these values at the fixed angle $\theta = 60$ degrees was desired in order to find the polarization of the light scattered at this angle by a cloud of drops in which various sizes are mixed. Fig. 5 shows an auxiliary graph made for this purpose. The plotted complex numbers are

$$x^{-1}S_1(\theta) = ix^{-1}i_1^*$$

and

$$x^{-1}S_2(\theta) = -ix^{-1}i_2^*.$$

Here i_1^* and i_2^* are the values tabulated by Gumprecht and the factor x^{-1} has been added in order to make the magnitude (except for phase and interference) constant for $x \rightarrow \infty$.

[3] R. O. Gumprecht, N. Sung, J. H. Chin, and C. M. Sliepcevich, *Jour. Opt. Soc. Amer.* vol. 42, p. 226; 1952.
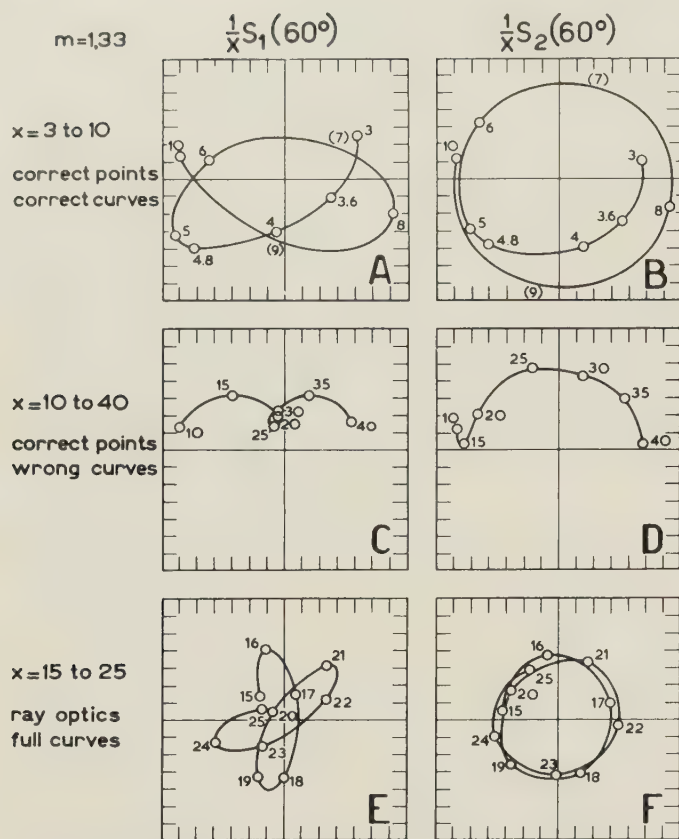
Fig. 5—Interpolation of the complex amplitudes of the radiation scattered by a sphere with $m = 1.33$ at $\theta = 60$ degrees in the two planes of polarization.

Fig. 6—Extinction curves for spheres with five real values of the refractive index. The scale of ρ is common to all curves.

In the upper part (A and B) of the figure, the points are sufficiently close to see how the curve runs, although more would be needed for a precise interpolation. The middle part (C and D) would seem all right to a person who has not been warned. He might draw the wrong conclusion that the scattering is strongly preponderant in polarization 2 for $\theta = 60$ degrees, $m = 1.33$ and x ranging from 20 to 30. The warning may come either from the upper curves for small x, or from the lower curves (E and F) that have been computed from geometrical optics. The latter drawings were constructed by adding the light refracted by the drop and the light externally reflected by it with the proper amplitude and phase. The period of the interference phenomenon from maximum to maximum amplitude is $\Delta x = 2.55$. The fact that $2\Delta x$ is close to the tabular interval 5, made it possible to draw fairly convincing curves in C and D.

The Extinction Curve

The extinction curve is the graph of Q_{ext} as a function of x for a body of a given form and composition and for a given state of polarization of the incident radiation. We shall omit the suffix. Theorems for arbitrary finite bodies and for infinitely long cylinders with arbitrary cross section show that Q is the real part of a factor $Q + iP$ occurring in the distant field expression for the forward scattered field.

The simplest case occurs for spheres or other circularly symmetric bodies with incident radiation propagating along the symmetry axis. We then may write the total distant field in the forward direction (on and near the positive z axis) as

$$u = e^{-ikz+i\omega t} + \frac{S(0)}{ikr} e^{-ikr+i\omega t}.$$

Here u may be any transverse component of the electric or magnetic field. We have .

$$Q + iP = \frac{4\pi}{k^2 G} S(0)$$

where G is the geometric shadow area.

Fig. 6 shows the extinction curves computed[4] by means of the Mie formulas for spheres with $m = 1.50$, $m = 1.33$, $m = 0.93$ and $m = 0.8$ and by means of the simple rigorous expression for $m = 1 \pm \epsilon$ (ϵ = small and real). The scale of $\rho = |2x(m - 1)|$ is common to all five curves.

Inspection of these curves shows that, except for the first ascending part where Q is still <1, the curves may be described as a superposition of four effects.

[4] $m = 1.33$: B. Goldberg, *Jour. Opt. Soc. Amer.*, vol. 43, p. 1221; 1953.

$m = 1.50$: unpublished results made available by Dr. R. Penndorf, AF Cambridge Res. Center, Cambridge, Mass.

$m = 0.80$ and 0.93: R. H. Boll, R. O. Gumprecht and C. M. Sliepcevich, *Jour. Opt. Soc. Amer.*, vol. 44, p. 18; 1954.

$m = 1 \pm \epsilon$: van de Hulst, *op. cit.*

1. A constant value $Q = 2$.

2. A big fluctuation with decreasing amplitude and maxima near $\rho = 4.1, 10.8, 17.2$ (period 2π).

3. A term that also goes to zero for $x \to \infty$ and lifts the average curve above 2 for $m > 1$ and depresses it below 2 for $m < 1$.

4. A ripple of a complex structure, that has its major period near $\Delta x = 0.80$. The latter statement appears more conclusive if also the other curves of Pendorf are inspected.

Our aim is to understand these components so fully that we can say if, and to what extent, similar effects are present in other directions (e.g. in radar cross sections) and for bodies of different shapes (e.g. cylinders, ellipsoids). We shall discuss each term consecutively.

1. The term $Q + iP = 2$ is found when Huygens' principle is applied to a plane-wave front with a part G missing. We shall call this component the *diffracted light*. That $Q = 2$ and not 1 can most readily be visualized by looking at the energy lost from the plane wave: the part incident on the body is lost anyway, and the part passing along it forms a diffraction pattern by which the same amount is lost again (and radiated at angles that differ only very little from the original direction).

2. The second component is due to the central beam that passes through the sphere along a diameter. We shall call this component the *refracted light*. If $m > 1$, the beam has a focus in or beyond the sphere and thereafter diverges. A reflection loss occurs upon incidence and emergence from the sphere and a phase jump π upon passing the focus. It is readily found from geometrical optics that the contribution of this term to $Q + iP$ is

$$\frac{1}{x} \frac{8m^2}{(m+1)\,|m^2-1|}\, e^{-2ix(m-1)-i\pi/2}.$$

Here m may be larger or smaller than 1. The major fluctuations in Fig. 6 are accurately rendered by this formula, both in amplitude and phase. A similar computation may be made for a beam that suffers two internal reflections and covers the diameter three times. Its contribution is found to be numerically insignificant compared to the effects 3) and 4) that will now be discussed.

We shall ascribe term 3) to grazing reflection near the "edge" of the sphere; i.e. near the curve that separates the illuminated and dark parts of the surface. We shall interpret term 4) as the contribution of the surface wave that has traveled 360 degrees around the body, taking several short-cuts through the body.

EDGE EFFECTS

External reflection gives light scattered in any direction. If this component is computed by means of geometrical optics, it is found that a finite nonvanishing amplitude results at $\theta = 0$, by grazing reflection. However, exactly in this limit the geometrical optics approximation fails to give the correct result. Scattering angles of the order of $x^{-1/3}$ are excepted from this approximation. This may be shown either by applying Huygens' principle to the wave-front formed by nearly grazing reflection, or from the limits imposed upon Debye's asymptotic transformation of the Mie series.

The latter derivation is briefly as follows. The "exceptional region" in the asymptotic formulas of the Bessel functions is characterized by the condition that $|p - x|$ is of the order of $x^{1/3}$. Here $p = n + \frac{1}{2}$ for spheres, $p = n$ for cylinders and $x = 2\pi a/\lambda$ as before. Terms with smaller p may be transformed by means of the main term of the nonexceptional expansion and give, after suitable separation, the ray-optics approximation by means of the method of stationary phase. It is then seen that terms of the order p correspond to rays incident at a distance $p\lambda/2\pi$ from the axis. So terms with $p < x$ correspond to light incident on the sphere, terms with $p > x$, with light passing along the sphere (there terms vanish indeed very rapidly) and terms with p near x, with waves that just hit or just miss the rounded edge of the body. The latter waves define a peculiar term in the complex amplitude of forward scattered light, that will be called the *edge term*.

The word edge should not be read as edge of a thin body but rather as the rounded edge of a thick body (e.g. edge of a spectrograph slit for light, or edge of an airplane wing for microwaves). Only the wave motion outside the body has to be considered at this stage, for the waves refracted into the body do not emerge in a manner to give stationary phase in the forward direction.

The edge terms probably have the following properties in all cases, including cylinders or spheres with arbitrary refractive index (perfect conductor, partial conductor or dielectric):

1. The phase difference with the diffracted light goes to a constant for $x \to \infty$, so the contribution to Q in Fig. 6 does not have a fluctuating character.

2. The magnitude (in $Q + iP$) is of the order of $x^{-2/3}$.

3. The phase factor is $\pm e^{-i\pi/3}$ for $x \to \infty$.

Attempts to derive the correct expressions for the edge terms by using the "exceptional expansions" of Debye have been unsuccessful so far. Two more advanced methods are available. One is the use of Watson's transformation proposed by Franz[5] as a modification of the earlier researches by van der Pol and Bremmer. This is a rigorous transformation but the method is limited to bodies of a simple geometric form for which a full solution is available to begin with. The second method is by means of an integral equation of the surface currents or fields at the surface. This method may be applied to bodies of arbitrary form and has first been used by Fock[6] and independently by Franz and Deppermann. The interrelation of the two methods was explained by Franz.

[5] W. Franz, *Z. Naturf.*, vol. 9a, p. 705; 1954.
[6] V. A. Fock, *Jour. Phys. USSR*, vol. 10, p. 130; 1946.

Barring, for a moment, the possibility that surface waves may creep all around the body, we see intuitively that the fields and currents near the "edge" should depend only on the state of polarization of the incident radiation, on the refractive index of the body (i.e. on the boundary conditions) and on the radius of curvature R of the body at the edge in a plane containing the direction of incidence and the normal. Fock has proved this by means of the integral equation and has computed the asymptotic form of the current distribution near the edge if kR is large. The body is a perfect conductor and the magnetic field of the incident wave is tangent to the surface at the edge. The current contains a factor $G(\xi)$, where $\xi = z(2R^2/k)^{-1/3}$ and z is the rectangular coordinate measured from the edge point inversely along the direction of propagation. The function G approaches 2 for $\xi \to \infty$ (on the illuminated side) and thus gives ordinary reflection. It goes to 0 for $\xi \to -\infty$ (on the dark side).

The edge term for perfectly conducting circular cylinders in Case II (H // axis) may be found from this function by numerical integration on the basis of Huygens' principle. In this way

$$Q + iP = 2 - (0.86 - 1.49\, i)\, x^{-2/3} + \ldots$$

was found in good agreement with the empirical formula derived from the results computed in the ordinary way. The empirical formula for perfectly conducting cylinders, Case I (E // axis) is

$$Q + iP = 2 + (1.00 - 1.73\, i)\, x^{-2/3} + \ldots\, .$$

It may be shown that the edge term for a sphere should be the sum of the edge terms for cylinders in Cases I and II. This would give

$$Q + iP = 2 + (0.14 - 0.24\, i)\, x^{-2/3} + \ldots\, .$$

In the later cases no precise checks on the numerical coefficients are available but the sign and the general order of magnitude of the coefficients is very plausible. The asymptotic point $Q + iP = 2$ is approached in Cases I and II from exactly opposite sides. This is due to the different boundary conditions that also give a different sign to the Fresnel reflection coefficients. In Case I the grazing reflection interferes destructively with the waves passing along the body. This means a higher extinction cross section. In Case II the sign is different and the extinction cross section is decreased. Perfectly conducting spheres show a combination of these effects and only a small positive excess is left.

It is very interesting to see what happens if the body is an imperfect conductor. The Fresnel reflection coefficients for perfect conductors are -1 in Case I and $+1$ in Case II. Those for any finite m both approach the value -1 for grazing angles. This means that the edge term in Case II for finite m should swing around to the same asymptote as holds for Case I. Fig. 7 shows the exact points of $Q + iP$ plotted for circular cylinders with
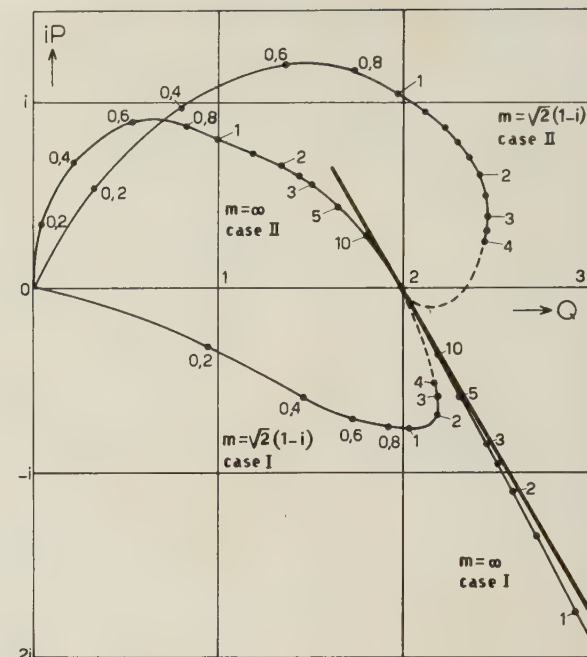
Fig. 7—Graphs of $Q+iP$ for circular cylinders for two values of the refractive index m. Case I: E // axis, Case II: H // axis. Values of $x = 2\pi a/\lambda$ are written with the computed points.

the refractive index ∞ (as described above) and with the refractive index $\sqrt{2}\,(1-i)$. The expected effect is indeed present.

After this digression on edge effects in general we may return to the interpretation of term (3) in the curves of Fig. 6. As soon as x is large enough the edge effects are determined essentially by grazing angles beyond the Brewster angle. This means equal sign of both reflection coefficients and destructive interference with the passing wave all along the edge of the sphere. Consequently, the extinction is considerably above $Q = 2$ (for $m > 1$). Adding the edge terms for perfectly conducting cylinders with a reversal in sign in Case II we find

$$(1.86 - 3.22\, i)\, x^{-2/3}$$

as the expected contribution (3) to the value of $Q + iP$ in the upper curves of Fig. 6. After taking out the main effects (1) and (2), the empirical coefficients $1.8 - 2.5\, i$ for $m = 1.33$ and 2.0 (only the real part) for $m = 1.20$ were found. This agreement is sufficiently close to show that the interpretation of term (3) as an edge term is qualitatively correct.

SURFACE WAVES

This subject can be discussed in a relatively concise manner as it will be treated in greater detail by Dr. Franz.

The asymptotic form of Fock's solution for the currents near the curved "edge" has the character of a damped wave running into the shadow region along the curved surface. The character of this wave is independent

of the injection mechanism, which, in this case, was the plane wave striking the curved edge at grazing angles. Mathematically, the surface wave is a solution of the homogeneous integral equation. Franz has shown that there are many more modes of the surface wave and that the edge phenomena, called here the injection mechanism, may be described as a superposition of many such modes. All are rapidly damped and only the least damped wave has practical significance at large distances from the edge.

For an examination of the character of these waves we shall look again at the perfectly conducting circular cylinder, illuminated by a plane wave that is plane-polarized with H // axis. Only this polarization gives a strong surface wave (set up at both edges). The exponent in the amplitude is

$$i\omega t - i x \varphi - (0.699 + 0.403 \, i) \, x^{1/3} \, \varphi \,,$$

where $x = ka$ and φ is the angular co-ordinate measured at the center. The *damping* is such that the amplitude reduces to $e^{-0.70 x^{1/3}}$ in completing one radian of the circumference. The energy lost by the wave is radiated at any point of the surface in tangential direction, i.e. in the direction in which the wave travels at that point. The *phase* is such that the wave travels with the velocity of light effectively at a distance $a(1 + 0.40 \, x^{1/3})$ from the center. So without having examined the detailed distribution of the fields outside the cylinder we may anticipate that the front of the surface wave has an effective width of the order of $x^{1/3} \, a = x^{4/3} \, . \, \lambda/2\pi$ outside the cylinder.

In the limit of zero curvature, i.e. $x \to \infty$, the wave becomes a full plane wave traveling alongside a perfectly conducting plane wall with H tangent to the surface. The damping per unit length becomes zero in this limit.

The same type of plane wave may be found as a special case of the Zenneck waves. These are plane damped waves in a nonabsorbing medium near the plane boundary with an absorbing medium (e.g. a metal). Again, H has to be tangent to the boundary. The damping occurs by Joule losses. If the metal is made perfectly conducting, the slant of the wave front vanishes and the damping vanishes.

It is plausible to assume that a surface wave with the same polarization can travel around a curved partial conductor and is damped by the two effects simultaneously: tangential radiation (spray) and Joule losses. A simple addition of the two damping effects may be sufficient in the first order. A wave equivalent to the plane undamped wave or to the Zenneck wave does not exist if the polarization is such that E is parallel to the surface. This explains at once why the surface waves traveling along a curved surface with this polarization are very weak.

Franz and Deppermann have shown convincingly that the surface waves just discussed are of major importance in explaining the radar cross sections of a perfectly conducting cylinder or sphere. The radiation (spray) of the surface wave after the completion of π radians along the surface interferes with the radiation centrally reflected from the body and gives large fluctuations with a period $\Delta x = 1.18$.

The spray in the forward direction, after completing 2π radians, is weaker as it has suffered double damping and has to interfere with the more powerful diffracted radiation. This explains why the influence of the surface wave is insignificant in the extinction cross sections of perfectly conducting cylinders and spheres. The expected period of the fluctuations is $\Delta x = 0.96$ near $x = 5$.

We may, finally, come to the explanation of term 4) in the extinction cross sections of partially transparent spheres; i.e. to the ripple in Fig. 6. Damping of the surface waves traveling along a plane boundary between two dielectrics occurs by continuous refraction into the denser medium in the direction prescribed by Snell's law for grazing incidence. Such waves have been studied, with a different injection mechanism, by Ott[7] and Maecker.[8]

The corresponding waves on a spherical surface, e.g. light waves on the surface of a raindrop, have not yet been studied in detail, although they are formally included in the work of Franz. There is one obvious but important difference with the waves on a plane surface. The energy lost by the main damping effect, namely radiation into the sphere, is bound to come back at some other part of the surface. It arrives there under the limiting angle of total reflection, i.e. precisely the most suitable angle to set up new surface waves.

The light thus travels a chord $2a \sin \alpha$ instead of an arc $2a\alpha$, where $\cos \alpha = 1/m$. This is a geometrical short-cut but a longer optical path by

$$2a (\tan \alpha - \alpha) = ha.$$

We may guess that the waves take the maximum number of "short-cuts," N, in completing the 360 degrees around the sphere, making the total optical path equal to $(2\pi + Nh) \, a$. The "ripple" may be interpreted as the interference of the waves radiated by these surface waves in the forward direction and the diffracted light. During the "short-cuts" the waves are protected from damping by spray, so their total loss in completing the circle may be considerably less than for perfectly conducting bodies.

The expected period of the ripple is

$$\Delta x = (1 + \frac{Nh}{2\pi})^{-1}.$$

The situation for $m = 1.5$ and $N = 3$ is sketched in Fig. 8. It gives $\Delta x = 0.79$. For $m = 1.333$, $N = 4$ we

[7] H. Ott, *Ann. der Phys.*, vol. 41, p. 443, 1942; vol. 4, p. 432, 1948.
[8] H. Maecker, *Ann. der Phys.*, vol. 4, p. 409; 1948.
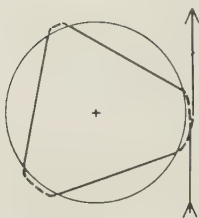
Fig. 8—Surface wave for $m = 1.5$ with three "short-cuts."

obtain $\Delta x = 0.83$. These values are in good agreement with the empirical periodicity. A certain amount of energy may skip a shortcut and remain at the surface until it can interfere with the emerging wave. The beat pattern set up by this interference must have a period $\Delta x = 2\pi/h$, which is 11.3 for $m = 1.5$ and 19.8 for $m = 1.333$. This checks very well with the empirical character of the ripple.

The backscatter from large water drops gives the natural phenomenon known as the glory, a series of colored and polarized rings around the shadow of the observer, or of an airplane, on the clouds. Although qualitatively understood,[9] there is still some doubt about its exact amount of polarization. It must be due to the light that has made two short-cuts and, in addition, has traveled 14 degrees of the circumference. An exact theory of the surface waves will also make the quantitative solution of this problem possible.

[9] H. C. van de Hulst, *Jour. Opt. Soc. Amer.*, vol. 37, p. 16; 1947.

Creeping Waves for Objects of Finite Conductivity

W. FRANZ† AND P. BECKMANN†

Summary—It is shown that it is not necessary to apply the van der Pol—Bremmer expansion in order to obtain the Watson residue series without remainder integral. There appear two kinds of residual waves. Those of the first kind do not enter the object and correspond to the usual creeping waves for objects of infinite conductivity. They arise from poles in the vicinity of the zeros of $H_v^{(1)}(ka)$. Residual waves of the second kind correspond to waves transversing the object and arise from poles in the vicinity of the zeros of $J_v(nka)$. They are of no importance in the case of strongly absorbing materials. Waves which are expected according to geometrical optics are obtained—as in the case of infinite conductivity— by splitting off an integral. Primary and reflected waves arise from two different saddle points of the same integrand which was thought of till now as only yielding the reflected waves. On the other hand the terms corresponding to the ingoing part of the primary wave give no contribution at all, but must be kept in order to assure the convergence of the integrals when shifting the path of integration.

I. GREEN'S FUNCTIONS FOR A CYLINDER OF FINITE CONDUCTIVITY

GREEN'S function for a cylinder with the radius a and the complex index of refraction n is given by the following series:

$$G(\rho_1, \rho_2, \varphi) = \frac{i}{8} \sum_0^\infty m \, \epsilon_m \cos m\varphi \, H_m^{(1)}(k\rho_2)$$
$$\{H_m^{(2)}(k\rho_1) - B_m H_m^{(1)}(k\rho_1)\} \qquad (1)$$

$$\epsilon_o = 1, \epsilon_m = 2 \text{ for } m > 0 \quad \rho_2 > \rho_1.$$

The coefficients B_m are determined by the boundary conditions. For E parallel to the axis of the cylinder we have:

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2)$$
$$B_m = \frac{H_m^{(2)\prime}(ka) \, J_m(\bar{k}a) - n \, H_m^{(2)}(ka) \, J_m{}'(\bar{k}a)}{H_m^{(1)\prime}(ka) \, J_m(\bar{k}a) - n \, H_m^{(1)}(ka) \, J_m{}'(\bar{k}a)}, \quad \bar{k} = nk;$$

analogous for H parallel:

$$B_m = \frac{H_m^{(2)\prime}(ka) \, J_m(\bar{k}a) - \dfrac{1}{n} H_m^{(2)}(ka) \, J'_m(\bar{k}a)}{H_m^{(1)\prime}(ka) \, J_m(\bar{k}a) - \dfrac{1}{n} H_m^{(1)}(ka) \, J'_m(\bar{k}a)}. \qquad (3)$$

Following Watson[1] we transform the original series (1) into an integral in the complex m-plane. We then deform the path of integration into a closed path around the poles of the integrand and evaluate the integral by means of a residue series. But, according to Imai[2] and Franz[3], for certain values of the angle we first have to split off an integral in order to obtain a good convergence of the residue series. It will be shown that the integral represents contributions which are expected according to geometrical optics.

Applying the well known method of van der Pol and Bremmer[4] to the case of the cylinder, in order to obtain a residue series without remainder integral, the coefficients B_m might be expanded into a series of ascending powers of $H_m^{(1)}(\bar{k}a) / H_m^{(2)}(\bar{k}a)$

$$B_m = \sum_0^\infty a_{m,l} \left[\frac{H_m^{(1)}(\bar{k}a)}{H_m^{(2)}(\bar{k}a)} \right]^l \qquad (4)$$

where the term containing

$$\left[\frac{H_m^{(1)}(\bar{k}a)}{H_m^{(2)}(\bar{k}a)} \right]^{l+1}, \qquad (5)$$

corresponds to waves which have been reflected inside the cylinder l-times. This procedure seems awkward mathematically since the expansion does not converge uniformly. As a consequence the results would contain residual waves (arising from poles in the vicinity of the zeros of $H_m^{(2)}(\bar{k}a)$) which are difficult to understand physically. We wish to show that a residue series without remainder integral can be obtained without expanding B_m.

II. TRANSFORMATION OF THE GREEN'S FUNCTION INTO A RESIDUE SERIES

Following Watson we transform the original series into an integral whose path of integration encloses the positive part of the real axis:

$$G = -\frac{1}{8} \int \frac{\cos v \, (\varphi - \pi)}{\sin v\pi} H_v^{(1)}(k\rho_2)$$
$$\{H_v^{(2)}(k\rho_1) - B_v H_v^{(1)}(k\rho_1)\} \, dv. \qquad (6)$$

We now deform the path of integration according to Fig. 1. The curve C_o encloses the poles of the integrand which are situated in the right half of the v-plane. The path C_1 is symmetrical with respect to the origin. C_o and C_1 are connected by arcs at infinity. We first have to show that these arcs give no contributions. To do this we give the following asymptotic formulas for the Hankel and Bessel functions for $v \gg z$ which may be obtained by the saddle point method following Debye.

The complex v-plane is divided into two regions. The boundary is given by the curves on which the zeros of $H_v(z)$ and $H_v{}'(z)$ are situated and the part of the real

† Munster University, Germany.
[1] G. N. Watson, "The diffraction of electric waves by the earth," *Proc. Roy. Soc. A*, vol. 95, pp. 83–99; 1919.
[2] I. Imai, "Beugung elektromagn. Wellen a.e. Kreiszylinder," *Z. Physik*, vol. 137, pp. 31–48; 1954, cf. sec. 2
[3] W. Franz, "Über die Greenschen Funktionen des Zylinders und der Kugel;" *Z. Naturf.* vol. 9a, pp. 705–716; 1954.
[4] H. Bremmer, "Terrestrial Radio Waves," Elsevier Publishing Co., New York; 1949.

Fig. 1

Fig. 2

Fig. 3
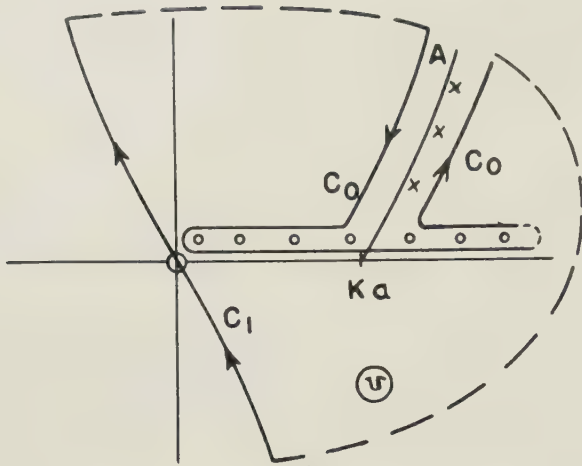
axis between z and $-z$. The curves on which the zeros are situated are denoted by h_1, h_{-1}, h_2, and h_{-2} (Figs. 2 and 3). For domain I in Fig. 2 we have

$$H_v^{(1)}(z) \sim \sqrt{\frac{2}{\pi v}} \left(\frac{ez}{2v} \right)^v ; \qquad (7)$$

for II,

$$H_v^{(1)}(z) \sim - i \sqrt{\frac{2}{\pi v}} \left(\frac{ez}{2v} \right)^{-v} ; \qquad (8)$$

while on the boundary the sum of both expressions has to be taken. Analogously is the Hankel function of the second kind for III (Fig. 3):

$$H_v^{(2)}(z) \sim - \sqrt{\frac{2}{\pi v}} \left(\frac{ez}{2v} \right)^v ; \qquad (9)$$

for IV,

$$H_v^{(2)}(z) \sim i \sqrt{\frac{2}{\pi v}} \left(\frac{ez}{2v} \right)^{-v} ; \qquad (10)$$

while on the boundary h again the sum of both representations has to be taken. To the right of h_1 and h_2 we have, for the Bessel function,

$$J_v(z) \sim \frac{1}{\sqrt{2 \pi v}} \left(\frac{ez}{2v} \right)^v , \qquad (11)$$

and to the left of h_{1-} and h_{-2},

$$J_{-v}(z) \sim \frac{1}{\sqrt{-2 \pi v}} \left(\frac{ez}{-2v} \right)^{-v} . \qquad (12)$$

In the center domain the Bessel function always equals half the larger of the two Hankel functions.

In (6) we form the difference $H_v^{(2)}(k \rho_1) - B_v H_v^{(1)}(k \rho_1)$. Considering the structure of the B_v we may in this expression substitute the Bessel functions for the Hankel functions of the second kind. Since $a < \rho_1 < \rho_2$, the increasing Hankel functions are always connected with more rapidly decreasing Bessel functions and the integrand vanishes for $0 < \varphi < \pi$ in the v-plane everywhere at infinity (except for the poles of the integrand which are located on curves extending to infinity; cf. Fig. 1). The free choice of J_v or $H_v^{(2)}$ in the expression containing
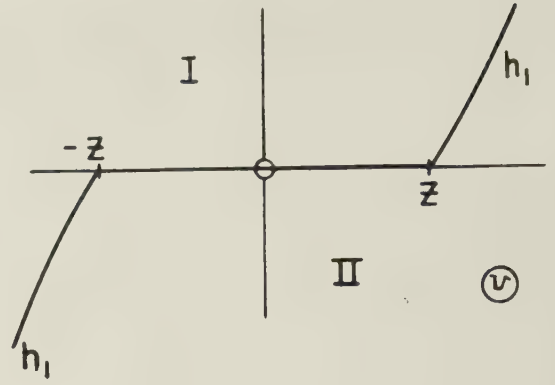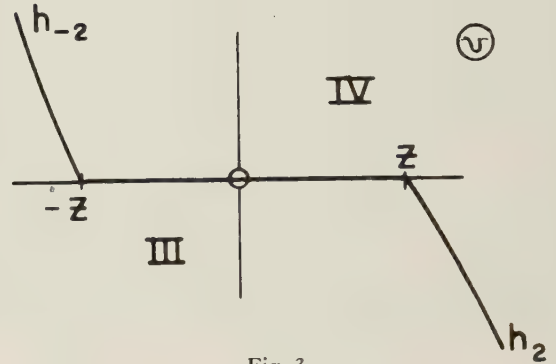
the primary wave allows one to shift path of integration freely in the v-plane (except for the poles). This is not possible if the primary wave is omitted.

Imai has done this in the case of a plane primary wave, but his derivations are wrong for two reasons: he is using the asymptotic formula (8) between h_1 and the imaginary axis, overlooking that the part DE of his path of integration is the arc on an infinitely small angle although it is of infinite length. Moreover even his formula makes his integral (3.6) diverge at the positive part of the imaginary axis since α is not a real number. So

$$| \alpha |^{-\alpha} = | \alpha |^{-\alpha_x} e^{\pi/2\alpha y} \to \infty \text{ for } \alpha \to i \infty.$$

A more thorough investigation shows that it is not possible to overcome the shortcomings of Imai's derivation by minor changes. For an angle of deflection smaller than $\frac{\pi}{2}$ there is no symmetrical path where (3.6) converges; for values $> \frac{\pi}{2}$ the final integral (3.15) is divergent everywhere at infinity in the lower half of the α-plane. In spite of this Imai's saddle point evaluation gives the correct result. The reason is—as we will see in section 5—that the correct path of integration leads over the same saddle point.

We now wish to deform C_1 into C_o. The present form of (6) does not allow this because of the denominator $\sin v\pi$. Let

$$F(v) = \cos v \ (\varphi - \pi) \ H_v^{(1)}(k \rho_2)$$
$$\{ \ H_v^{(2)}(k \rho_1) - B_v \ H_v^{(1)}(k \rho_1) \ \}. \qquad (13)$$

Since C_1 is symmetrical with respect to the origin we may put

$$\int_{C_1} \frac{F(v)}{\sin v\pi}\, dv = \frac{1}{2} \int_{C_1} \frac{F(v) - F(-v)}{\sin v\pi}\, dv. \qquad (14)$$

The difference $F(v) - F(-v)$ contains a factor

$$J_v(\bar{k}a)\, J_{-v}{}'(\bar{k}a) - J_v{}'(\bar{k}a)\, J_{-v}(\bar{k}a) = -\frac{2\sin v\pi}{\pi \bar{k}a}, \qquad (15)$$

compensating the denominator $\sin v\pi$. We now may deform C_1 into C_o and obtain, after changing the direction of the path C_1,

$$G = \frac{-1}{16} \int_{C_1} \frac{\cos v\,(\varphi-\pi)}{\sin v\pi} H_v{}^{(1)}(k\rho_2) \Big\{ 2H_v{}^{(2)}(k\rho_1) \\ - H_v{}^{(1)}(k\rho_1)\,(B_v + e^{2\pi i v}\, B_{-v}) \Big\}\, dv. \qquad (16)$$

This integral may now be evaluated by a residue series. The poles may be divided into two groups: the first one containing those in the vicinity of the zeros of $H_v{}^{(1)}(ka)$ and $H_v{}^{(1)}{}'(ka)$, the second one including those in the vicinity of the zeros of $J_v(\bar{k}a)$, $J_v{}'(\bar{k}a)$, $J_{-v}(\bar{k}a)$, and $J'_{-v}(\bar{k}a)$. We shall see that the latter may be neglected in the case of cylinders of strongly absorbing materials.

III. Evaluation of the Residue Series for Strongly Absorbing Materials

We observe that B_v may be written as[5]

$$B_v = \frac{H_v{}^{(2)}(ka)}{H_v{}^{(1)}(ka)} \left\{ \frac{ln'\, H_v{}^{(2)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a)}{ln'\, H_v{}^{(1)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a)} \right. \qquad (17)$$
$$- \frac{H_v{}^{(1)}(\bar{k}a)\,(ln'\, H_v{}^{(1)}(ka) - ln'\, H_v{}^{(2)}(ka))}{H_v{}^{(2)}(\bar{k}a)\,(ln'\, H_v{}^{(1)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a))^2} \times$$
$$\frac{n\,(ln'\, H_v{}^{(1)}(\bar{k}a) - ln'\, H_v{}^{(2)}(\bar{k}a))}{\left\{ 1 + \dfrac{H_v{}^{(1)}(\bar{k}a)}{H_v{}^{(2)}(\bar{k}a)} \cdot \dfrac{ln'\, H_v{}^{(1)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(1)}(\bar{k}a)}{ln'\, H_v{}^{(1)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a)} \right\}} \left. \right\}.$$

We wish to show that the residues arising from the second term in (17) may be neglected if the complex index of refraction has a great imaginary part. We observe that the numerator of the second term in (17) contains the Wronskian of $H_v{}^{(1)}(\bar{k}a)$ and $H_v{}^{(2)}(\bar{k}a)$. The order of magnitude is therefore determined by the factor $1/[H_v{}^{(2)}(\bar{k}a)]^2$ which is small for a positive real part of v (this does not hold for the vicinity of the zeros of $H_v{}^{(2)}(\bar{k}a)$. There each of the terms in (17) has poles. But B, as before, is regular there. So we do not have to consider any residues.)
Accordingly, we have

$$B_v \simeq -\frac{H_v{}^{(2)}(ka)}{H_v{}^{(1)}(ka)}\, R_{12};$$

$$R_{12} = -\frac{ln'\, H_v{}^{(2)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a)}{ln'\, H_v{}^{(1)}(ka) - n^{\pm 1}\, ln'\, H_v{}^{(2)}(\bar{k}a)}. \qquad (18)$$

In the limit $a \to \infty$, R_{12} is the well-known reflection coefficient of Fresnel. Eq. (16) now reads

[5] $ln'\, H_v(z)$ stands for $H_v{}'(z) / H_v(z)$.

$$G \simeq -\frac{1}{8} \int_{C_o} \frac{\cos v(\varphi-\pi)}{\sin v\pi} H_v{}^{(1)}(k\rho_2) \Big\{ H_v{}^{(2)}(k\rho_1) \\ + H_v{}^{(1)}(k\rho_1) \frac{H_v{}^{(2)}(ka)}{H_v{}^{(1)}(ka)} R_{12} \Big\}\, dv. \qquad (19)$$

If with v_l we denote the zeros of the denominator of R_{12} in the vicinity of the curve A (cf. Fig. 1) we obtain the residue series

$$G \simeq \frac{i\pi}{4} \sum_{1}^{\infty} l\, \frac{\cos v_l(\varphi-\pi)}{\sin v_l\pi}\, H_{vl}{}^{(1)}(k\rho_2)\, H_{vl}{}^{(1)}(k\rho_1)$$
$$\times \frac{H_{vl}{}^{(2)}(ka)}{H_{vl}{}^{(1)}(ka)}\, \frac{ln'\, H_{vl}{}^{(2)}(ka) - n^{\pm 1}\, ln'\, H_{vl}{}^{(2)}(\bar{k}a)}{\dfrac{\partial}{\partial v}\Big(ln'\, H_v{}^{(1)}(ka) - n^{\pm 1} ln'\, H_v{}^{(2)}(\bar{k}a) \Big)_{v=v_l}}, \qquad (20)$$

which may be written using the Wronskian of $H_{vl}{}^{(1)}(ka)$ and $H_{vl}{}^{(2)}(ka)$;

$$G \simeq \frac{1}{ka} \sum_{1}^{\infty} l\, \frac{\cos v_l(\varphi-\pi)}{\sin v_l\pi}\, H_{vl}{}^{(1)}(k\rho_2)\, H_{vl}{}^{(1)}(k\rho_1)$$
$$\times \frac{1}{[H_{vl}{}^{(1)}(ka)]^2}\, \frac{1}{\dfrac{\partial}{\partial v}\Big(ln'\, H_v{}^{(1)}(ka) - n^{+1} ln'\, H_v{}^{(2)}(\bar{k}a) \Big)} \qquad (21)$$
$$v=v_l$$

This expression represents creeping waves as in the case of a cylinder of infinite conductivity, but with different damping and amplitude. In order to discuss the influence of the radius a and the index of refraction it is necessary to determine the values of the v_l.

IV. Discussion of the Damping Factor for the Creeping Waves

The damping factor of the creeping waves is determined by the zeros of

$$\frac{H_v{}^{(1)}{}'(ka)}{H_v{}^{(1)}(ka)} - n^{\pm 1} \frac{H_v{}^{(2)}{}'(\bar{k}a)}{H_v{}^{(2)}(\bar{k}a)}; \qquad (22)$$

n^{+1} is valid for E parallel to the axis, n^{-1} is for H parallel. We denote these two cases by the subscripts π and σ. Using the asymptotic formulas given by Debye we obtain

$$\frac{H_v{}^{(2)}{}'(\bar{k}a)}{H_v{}^{(2)}(\bar{k}a)} \sim -i\sqrt{1 - \frac{v^2}{\bar{k}^2 a^2}}. \qquad (23)$$

Since only the zeros in the vicinity of $v=ka$ are of importance we may put $v=ka$ in this expression and obtain

$$\frac{H_v{}^{(2)}{}'(\bar{k}a)}{H_v{}^{(2)}(\bar{k}a)} \sim -i\sqrt{1 - \frac{1}{n^2}} = -i\,r. \qquad (24)$$

For $H_v{}^{(1)}(ka)$ and its derivative we substitute the Airy integral and its derivative, using the notation of Franz[3],

$$\frac{H_v{}^{(1)}{}'(ka)}{H_v{}^{(1)}(ka)} \sim -y\, \frac{A'(q)}{A(q)}, \qquad (25)$$

with

$$y = e^{-i\pi/3} \Big(\frac{6}{ka} \Big)^{1/3}; \quad q = y\,(v-ka). \qquad (26)$$
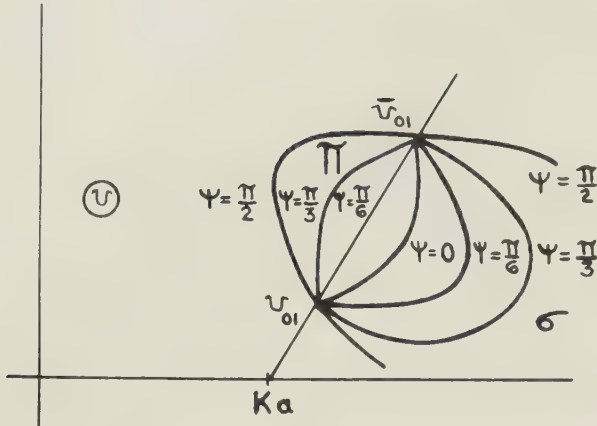
Fig. 4

The zeros are now determined by

$$\frac{A\,(q_l)}{A'\,(q_l)} = \frac{1}{r}\,n^{\pm 1}\,e^{-i\pi/2}\,y, \tag{27}$$

with

$$v_l = ka + \frac{q_l}{y}. \tag{28}$$

According to (27) we see that for increasing $|n|$ and ka the v_π move from the zeros of $A'(q)$ [resp. $H_v^{(1)'}(ka)$] towards the zeros of $A\,(q)$ [resp. $H_v^{(1)}(ka)$] along a curve C_π, while the v_σ for increasing $|n|$ and decreasing ka move from the zeros of $A\,(q)$ [resp. $H_v^{(1)}(ka)$] towards the zeros of $A'(q)$ [resp. $H_v^{(1)'}(ka)$]. If we expand $A\,(q)$ and $A'\,(q)$ in the vicinity of the zeros of $A\,(q)$ [$A\,(q_o)=0$] and $A'\,(q)$ [$A'\,(q_o)=0$] we have, with $q=q_o+p$,

$$p = \frac{1}{r}\,|n|^{\pm 1}\,|y|\,e^{-i(5/6\pi \pm \psi)}\;;\; n = |n|\,e^{i\psi}; \tag{29}$$

and with $q=q_o+p$, using the differential equation of the Airy integral, i.e.

$$A'' = -\frac{q}{3}\,A, \tag{30}$$

we obtain

$$p = r\,|n|^{\pm 1}\,\frac{3}{q_o\,|y|}\,e^{-i(\pi/6 \mp \psi)}. \tag{31}$$

We see from (30) and (31) as already from (27) that for real values of n the curves C_π and C_σ coincide. For $\psi > 0$ they start at q_o and q_o symmetrically with respect to the curve for $\psi = 0$ (Fig. 4). For ka large, i.e., in the vicinity of q_o, the damping increases with increasing ψ for π and σ. For ka small, i.e., in the vicinity of q_o, for σ the damping decreases with increasing ψ. For $\psi = \frac{\pi}{6}$ the damping is equal to the damping of a cylinder of infinite conductivity. For $\psi > \frac{\pi}{6}$ the damping becomes smaller. For π the damping first increases with increasing ψ and has for $\psi = \frac{\pi}{3}$ a maximum.

For actual computation of the v_l, (27) is not exact enough—we may not neglect terms proportional to $(ka)^{-2/3}$ and obtain (cf. Franz and Galle[6])

[6] W. Franz and R. Galle, "Semiasymptotische Reihen für die Beugung einer ebenen Welle a. Zylinder;" Z. Naturf. vol. 10a, pp. 374-378; 1955.

$$\frac{H_{vl}^{(1)}\,(ka)}{H_{vl}^{(1)'}\,(ka)} \sim -\frac{1}{y}\,\frac{A+y^2\,\dfrac{q_l}{45}\,A+y^2\,\dfrac{q_l^2}{180}\,A'}{A'-y^2\left(\dfrac{1}{30}+\dfrac{q_l^3}{540}\right)A-\dfrac{y^2 q_l}{45}\,A'}. \tag{32}$$

With the abbreviation

$$\delta = \frac{1}{r}\,y\,n^{\mp 1}\,e^{-i\pi/2}, \tag{33}$$

our final equation becomes

$$\frac{A\,(q_l)}{A'\,(q_l)} = \delta\,\frac{1 - \dfrac{y^2 q_l}{45} - \dfrac{y^2 q_l^2}{180\,\delta}}{1 + \dfrac{y^2 q_l}{45} + y^2\delta\left(\dfrac{1}{30}+\dfrac{q_l^3}{540}\right)}. \tag{34}$$

For steel, Fig. 5 shows $Im(v_1)$ as a function of ka. For reference, experimental values given by Hartnagel are included.
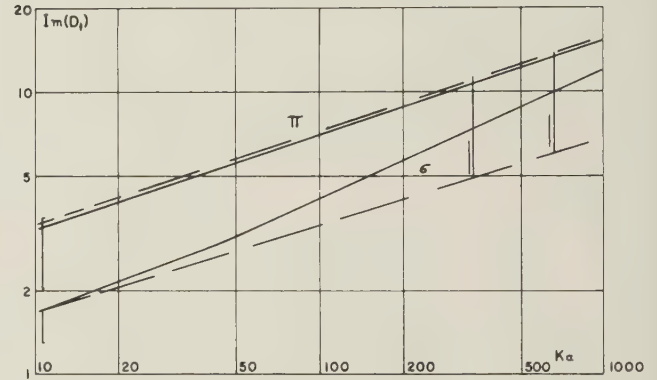
Fig. 5—Solid curves: steel; broken curves: infinite conductivity; [experimental values of diffraction of light by razor blades (Hartnagel).]

V. DISCUSSION OF THE REFLECTED AND PRIMARY WAVE

As to the convergence of the residue series (21) the same considerations as in the case of a cylinder of infinite conductivity are valid.[3] In the region which is reached by the primary wave we have to split off an integral with the aid of the identity

$$\cos v(\varphi - \pi) = e^{iv\pi}\cos v\varphi - i\,e^{iv\varphi}\sin v\pi. \tag{35}$$

We obtain a residue series with slightly changed trigonometric functions:

$$G_{cr} = \frac{1}{ka}\sum_l^\infty l\,\frac{\cos v_l\varphi}{\sin v_l\pi}\,e^{iv_l l\pi}\,H_{vl}^{(1)}(k\rho_2)\,H_{vl}^{(1)}(k\rho_1)$$

$$\times\,\frac{1}{[H_{vl}^{(1)}\,(ka)]^2}\frac{1}{\dfrac{\partial}{\partial v}\left(ln'\,H_v^{(1)}(ka)-n^{\pm 1}\,ln'\,H_v^{(2)}(\bar{k}a)\right)_{v=v_l}} \tag{36}$$

and in addition the integral

$$G_{p,r} \approx \frac{i}{8}\int_D e^{iv\varphi}\,H_v^{(1)}(k\rho_2)\,H_v^{(1)}(k\rho_1)\,\times$$

$$\frac{H_v^{(2)}(ka)}{H_v^{(1)}(ka)}\,R_{12}\,dv. \tag{37}$$

On both sides of the curve on which the poles are located the integrand becomes small at infinity. In the whole lower half of the v-plane the integrand increases—
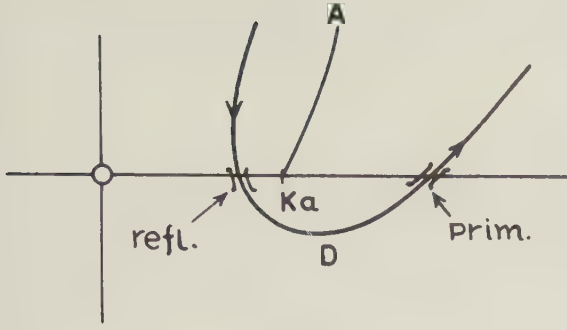
Fig. 6

with the exception of a small region around each of the zeros of $H_v^{(2)}(ka)$. We put the path of integration as shown in Fig. 6. It leads from the valley left of A to the first zero of $H_v^{(2)}(ka)$ and from there into the valley right of A, passing two saddle points (cf. section III). In order to determine the saddle point left of ka we may use the Debye asymptotic formulas, and obtain

$$\frac{\partial}{\partial v}\left(v\varphi + \sqrt{k^2\rho_1^2 - v^2} - v\arccos\frac{v}{k\rho_1} + \sqrt{k^2\rho_2^2 - v^2}\right.$$

$$\left. - v\arccos\frac{v}{k\rho_2} - 2\sqrt{k^2a^2 - v^2} + 2v\arccos\frac{v}{ka}\right) = 0 \quad (38)$$

at the saddle point. It follows that

$$\arccos\frac{v_s}{k\rho_1} + \arccos\frac{v_s}{k\rho_2} = 2\arccos\frac{v_s}{ka} + \varphi. \quad (39)$$
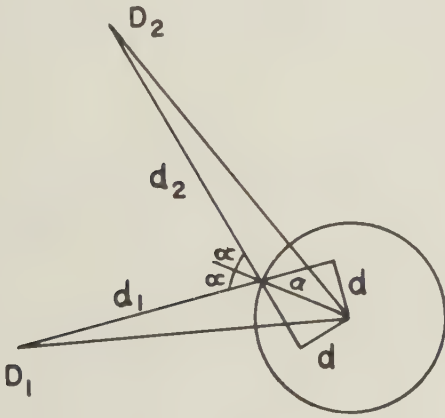
Fig. 7

This equation has a simple geometrical meaning as may be seen in Fig. 7. Let us draw the ray which is reflected from P_1 towards P_2 according to geometrical optics. If we denote its distance from the center by d and the angle of reflection by α we have

$$d = a\cos\left(\frac{\pi}{2} - \alpha\right) \quad (40)$$

and kd satisfies equation (39) if substituted for v_s. If we introduce the two chords of the reflected ray as abbreviations;

$$d_1 = \sqrt{\rho_1^2 - d^2} - \sqrt{a^2 - d^2},$$
$$d_2 = \sqrt{\rho_2^2 - d^2} - \sqrt{a^2 - d^2}, \quad (41)$$

we obtain the asymptotic formula

$$G_{refl} \sim \frac{i\,e^{-i\pi/4}}{4\sqrt{\pi k}}\sqrt{\frac{2\,(a^2 - d^2)^{1/2}}{2\,d_1d_2 + (a^2 - d^2)^{1/2}\,(d_1 + d_2)}}\times$$

$$R_{12s}\,e^{ik(d_1+d_2)} \quad (42)$$

The same expression for the reflected wave is obtained from geometrical optics.

At the right saddle point of (37) we have $H_v^{(1)}(ka) \sim - H_v^{(2)}(ka)$. Therefore the position of the saddle point is determined by

$$\arccos\frac{v_s}{k\rho_1} + \arccos\frac{v_s}{k\rho_2} = \varphi. \quad (43)$$

As may be seen in Fig. 8, v_s/k is the distance of the ray connecting P_1 and P_2 from the center. If we use the proper—i.e., positive— sign of the arccosines in (43), only Q between P_1 and P_2 is allowed. If φ decreases and reaches the value arccosine ρ_1/ρ_2 the saddle point approaches $k\rho_1$. Upon further decreasing of φ the saddle point approaches the first zero of $H_v^{(1)}(k\rho_1)$. In this case the path of integration connects the zeros of $H_v^{(1)}(k\rho_1)$.
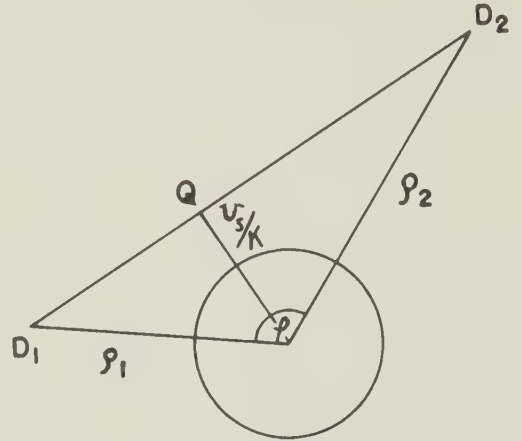
Fig. 8

We therefore have an infinite number of saddle points which are located between the zeros of $H_v^{(1)}(k\rho_1)$.

For $\varphi > \arccos\rho_1/\rho_2$ a single saddle point gives the primary wave. In order to show that the saddle point(s) to the right of ka always give(s) the primary wave we give an integral representation of the primary wave.

VI. Integral Representation of the Primary Wave

Applying the Watson transformation to the well-known series expansion of the Green's function of free space, we have

$$G_o = \frac{1}{4}\int\frac{\cos v\,(\varphi - \pi)}{\sin v\pi}H_v^{(1)}(k\rho_2)\,J_v(k\rho_1)\,dv, \quad \rho_1 < \rho_2, \quad (44)$$

which may be written as

$$G_o = \frac{1}{8}\int\frac{H_v^{(1)}(k\rho_2)}{\sin v\pi}\left\{e^{iv(\varphi-\pi)}\,J_v(k\rho_1) + e^{iv(\pi-\varphi)}\right.$$

$$\left. J_v(k\rho_1)\right\}\,dv. \quad (45)$$

The path of integration may be deformed into the curve on which the zeros of $H_v^{(1)}(k\rho_1)$ are located if $0 < \varphi < \pi$. Because the path of integration now is symmetrical with respect to the origin we may substitute $-v$ for v in the second term of (45). Using

$$J_v(z) = \frac{1}{2} \left(H_v^{(1)}(z) + H_v^{(2)}(z) \right) \qquad (46)$$

and

$$H_v^{(1)}(z) = e^{-iv\pi} H_{-v}^{(1)}(z) \; ; H_v^{(2)}(z) = e^{iv\pi} H_{-v}^{(2)}(z), \qquad (47)$$

we obtain

$$G_o = \frac{i}{8} \int e^{iv\varphi} H_v^{(1)}(k\rho_2) H_v^{(1)}(k\rho_1) \, dv. \qquad (48)$$

The integrand is the same as in (37) for $v > ka$. Although the path is not exactly the same, in both cases it leads over the saddle points located in the right half of the v-plane. So we have shown that the integral (37) always gives the primary and reflected wave.

It may be surprising to obtain both reflected *and* primary wave from the second term of (1). This is contrary to the interpretation of the first term of (1) as given by van der Pol and Bremmer. Bremmer, on page 33 of his book, expects—in the case of a sphere which is quite analogous—the first term of (1) "to operate as primary field," while we have seen that in the transformed integral it gives no contribution at all. Moreover it is easy to see that the first term may not be split off at all since the series would diverge. (This holds for both series in (6) on page 33 of Bremmer's book.)

VII. APPLICATION TO THE GREEN'S DYADIC OF THE SPHERE

While we have treated the cylinder as a purely two dimensional problem by means of the scalar wave equation, the sphere case as an entirely three dimensional problem leads to vector wave functions satisfying Maxwell's equations. Although the general formulas in this case are somewhat more complicated the Watson transformation may be performed quite analogous to the cylinder. The general solution may be expressed by a Green's dyadic composed of two series of the form

$$\Gamma(r,r') = \sum (2m+1) \Lambda H_{m+\frac{1}{2}}^{(1)}(kr') \left\{ H_{m+\frac{1}{2}}^{(2)}(kr) - B_{m+\frac{1}{2}} H_{m+\frac{1}{2}}^{(1)}(kr) \right\} P_m(\cos \vartheta), \; r' > r \qquad (49)$$

ϑ is the angle between r and r', Λ a differential operator which is an even function of $n+\frac{1}{2}$. There appear two kinds of coefficients

$$B_v = \cfrac{H_v^{(2)}(ka)[J_v'(\bar{k}a) + \frac{1}{2\bar{k}a} J_v(\bar{k}a)] - nJ_v(\bar{k}a)}{H_v^{(1)}(ka)[J_v'(\bar{k}a) + \frac{1}{2\bar{k}a} J_v(\bar{k}a)] - nJ_v(\bar{k}a)}$$

$$\cfrac{[H_v^{(2)'}(ka) + \frac{1}{2ka} H_v^{(2)}(ka)]}{[H_v^{(1)'}(ka) + \frac{1}{2ka} H_v^{(1)}(ka)]} \qquad (50)$$

B_v : *substitute $1/n$ for n in* (50). (50a)

Their structure allows one as in the case of the cylinder to substitute J_v for $H_v^{(2)}$, so the path of integration may be shifted in the same way. After performing the first step of the Watson transformation we obtain integrals of the form:

$$\Gamma = i \int \frac{v}{\cos v\pi} \Lambda H_v^{(1)}(kr') \left\{ H_v^{(2)}(kr) - B_v H_v^{(1)}(kr) \right\} P_v(-\cos \vartheta) \, dv. \qquad (51)$$

When symmetrizing the integrand of path C_1 there appears a factor $\sin v\pi$, as in the case of the cylinder, which now does not compensate the denominator $\cos v\pi$. So there are still poles on the real axis which do not allow transforming the remainder integral into a residue series of the Watson type. Here a specific feature of the spherical harmonics helps. After symmetrizing, the integral is of the form

$$\int \frac{v \sin v\pi}{\cos v\pi} F(v^2) P_{v-\frac{1}{2}}(-\cos \vartheta) \, dv,$$

where $F(v^2)$ is an even function of the variable v containing differential operators. The integrand here may be written as follows:

$$\frac{vF(v^2)}{2\cos v\pi} \left[e^{i(v-\frac{1}{2})\pi} P_{v-\frac{1}{2}}(-\cos \vartheta) - P_{v-\frac{1}{2}}(\cos \vartheta) \right].$$

$$-\frac{vF(v^2)}{2\cos v\pi} \left[e^{-i(v+\frac{1}{2})\pi} P_{v-\frac{1}{2}}(-\cos \vartheta) - P_{v-\frac{1}{2}}(\cos \vartheta) \right].$$

Using the symmetry of the path C_1 in the second term $-v$ may be substituted for v. So the integral becomes:

$$\int \frac{vF(v^2)}{\cos v\pi} \left[e^{i(v-\frac{1}{2})\pi} P_{v-\frac{1}{2}}(-\cos \vartheta) - P_{v-\frac{1}{2}}(\cos \vartheta) \right] dv. \qquad (52)$$

Here the square bracket is zero for all positive half integer values of v; so we have no more poles on the real axis. Consequently the remainder integral may be shifted to the path C_o (Fig. 1) and we get for the total Green's dyadic two integrals of the form

$$\Gamma = -\int \frac{v \, dv}{\sin(2v\pi)} \Lambda(v^2) H_v^{(1)}(kr') \left[(e^{3\pi iv} B_{-v} - e^{-i\pi v} B_v) P_{v-\frac{1}{2}}(\cos \vartheta) + i(B_v - e^{2\pi iv} B_{-v}) P_{v-\frac{1}{2}}(\cos \vartheta) \right] H_v^{(1)}(kr). \qquad (53)$$

This now may be transformed into a residues series. For strongly absorbing materials $e^{2\pi iv} B_{-v}$ may safely be approximated by B_v so the terms with $P_{v-\frac{1}{2}}(\cos \vartheta)$ vanish and a factor $\sin v\pi$ of the denominator is cancelled. The result is much like the one in case of an ideally reflecting sphere as treated by Franz[3] previously; just as in that paper, in order to obtain the geometrical waves in the lit part of space an integral has to be split off, introducing spherical harmonics of the second kind (they already appear in (52)). The saddle point evaluation of this integral yields as in the case of the cylinder the reflected wave from a saddle point to the left of ka and the primary wave from a saddle point to the right of ka. Here again the B_v terms of (49) give both reflected and primary wave.

A Method for the Asymptotic Solution of Diffraction Problems

R. TIMMAN†

Summary—The equation for the propagation of harmonic waves in a homogeneous medium is considered as the transform of an hyperbolic equation in one more variable. The boundary value problem of diffraction theory can, by this Laplace transform, be related to Cauchy's problem. The transformed problems are solved for 2 + 1 variables by methods introduced by Evvard and Ward in supersonic airfoil theory. As an example the diffraction problem for a strip is worked out and an asymptotic expression for the transmission cross section is given.

I. INTRODUCTION

THIS PAPER is a condensed and revised version of the thesis worked out under the author's direction by A. P. Burger.[1] We consider the problem of acoustic diffraction by a plane perfectly reflecting screen S in a finite part of the xz plane of a Cartesian system. If the time dependence is given by the factor $e^{-i\nu t}$, the total velocity potential satisfies the equation

$$\varphi_{xx} + \varphi_{yy} + \varphi_{zz} + k^2\varphi = 0, \qquad (1)$$

where $k = \nu/c$ and c is the sound velocity. For a perfectly reflecting screen the boundary conditions are

$$\frac{\partial \varphi}{\partial n} = 0 \text{ on } S.$$

If the incident wave has the normal $n = (n_1, n_2, n_3)$

$$\varphi^i = e^{ik(r,n)} \qquad (2)$$

we put

$$\varphi^{tot} = e^{ik(r,n)} + \varphi, \qquad (3)$$

where φ is the disturbance potential, which must satisfy in infinity Sommerfeld's condition. The boundary condition on S is

$$\varphi_y = ikn_2 . e^{ik(r,n)}. \qquad (4)$$

On both sides of S, φ_y has the same value and φ_y is an even function of y. Hence φ is an odd function, which involves that $\varphi = 0$ on S', the complement of S in the xz plane, where φ must be continuous. On the edge of S, φ must be finite.

The boundary value problem for φ, with conditions

$$\varphi_y = \frac{1}{ik} f(x, z) \quad \text{on } S \qquad (5)$$

$$\varphi = 0 \qquad \text{on } S'$$

together with edge and radiation conditions is not solved directly but transformed into another boundary value problem in a four-dimensional space with coordinates x, y, z, and a time variable t. Since (1) results from the four-dimensional wave equation

$$\psi_{xx} + \psi_{yy} + \psi_{zz} - \psi_{tt} = 0, \qquad (6)$$

by a one-sided Fourier transform (or Laplace transform),

$$\varphi(x, y, z; k) = \int_0^\infty e^{ikt}\psi(x, y, z, t) \, dt, \qquad (7)$$

the problem is transformed into the following boundary value problem for ψ

$$\begin{aligned}
\psi &= 0 & & t \le 0 \\
\psi_y &= f(x, z) & y = 0, \quad t > 0 & \quad \text{on } S \\
\psi &= 0 & y = 0, \quad t > 0 & \quad \text{on } S';
\end{aligned}$$

which is closely related to Cauchy's problem for hyperbolic equations. Problems of this kind are well-known in the theory of airfoils moving at supersonic speeds in three-dimensional space and hence the results of this theory are only directly applicable to two-dimensional diffraction problems. We consider henceforth only the problem of diffraction by a strip, or according to Babinet's principle, diffraction by a slit.

II. THE PROBLEM OF DIFFRACTION BY A SLIT

The slit extends along the lines $y = 0$, $|x| = 1$, and all quantities are independent of z. The transformed problem is

$$\psi_{xx} + \psi_{yy} - \psi_{tt} = 0 \qquad (8)$$

$$\begin{aligned}
\psi_y &= f(x) & y = 0, |x| < 1 & \quad t > 0, \\
\psi &= 0 & y = 0, |x| > 1 & \quad t > 0, \quad (9) \\
\psi &= \psi_y = 0 & & t \le 0.
\end{aligned}$$

The solution of this problem is based upon the Hadamard-Riesz theory of Cauchy's problem.[2,3]

If

$$\frac{1}{r} = \frac{1}{\sqrt{(t - r)^2 - (x - \xi)^2 - (y - \eta)^2}}$$

is the elementary solution with respect to a point $P(x, y, t)$ only inside the forward characteristic cone of P, then for a surface S, which closes this characteristic cone, Hadamard has derived the theorem

$$\psi(x, y, t) = \frac{1}{2\pi} \left| \iint_S \left(\frac{\partial \psi}{\partial \nu} \frac{1}{r} - \psi \frac{\partial}{\partial \nu} \frac{1}{r} \right) ds. \qquad (10)$$

Here ν is the "conormal" of S, which has direction cosines $(n_1, n_2, -n_3)$ instead of (n_1, n_2, n_3) of the normal.

If both ψ and ψ_n are given on S, this formula directly solves Cauchy's problem, but in our case it does not give this simple result. We apply the theorem to the region formed by the characteristic cone (Fig. 1); the

[1] A. P. Burger, "On the Asymptotic Solution of Wave Propagation and Oscillation Problems." Thesis, Delft, 1955; Rep. F 157, National Huchtvaart Laboratorium, Amsterdam.

[2] J. Hadamard, Lectures on Cauchy's Problem, Yale; 1923.
[3] M. Riesz; *Act. Math.*, vol. 81, pp. 1–223; January, 1949.

part of the plane $y = 0$ inside this cone and the plane $t = 0$. Here $\psi = \psi_y = 0$ and on the (x, t) plane

$$\frac{\partial}{\partial \nu} = -\frac{\partial}{\partial \eta}.$$

Hence we obtain

$$\psi(x, y, t) = -\frac{1}{2\pi} \left| \iint_D \left(\frac{\partial \psi_+}{\partial \eta} \frac{1}{r} - \psi_+ \frac{\partial}{\partial \eta} \frac{1}{r} \right) d\xi \, d_\tau, \quad (11)$$

where the $+$ signs apply to the value at the upper side of the plane. The corresponding theorem for the part of the cone below the plane $y = 0$ yields the equation

$$0 = \frac{1}{2\pi} \left| \iint_D \left(\frac{\partial \psi_-}{\partial \eta} \frac{1}{r} - \psi_- \frac{\partial}{\partial \eta} \frac{1}{r} \right) d\xi \, d_\tau. \quad (12)$$
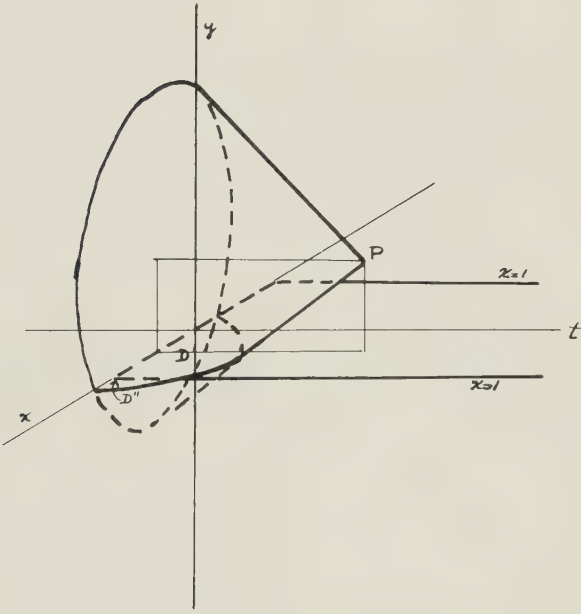
Fig. 1

Since ψ is an odd function of y we can eliminate ψ and we obtain by subtraction

$$\psi(x, y, t) = \frac{1}{\pi} \iint_D \frac{\psi_\eta(\xi, 0, \tau) 1 \, d\xi \, d_\tau}{\sqrt{(t - \tau)^2 - (x - \xi)^2 - (y^2)}}. \quad (13)$$

(For the complementary boundary value problem we add).

The value of ψ_η is given on the strip $|\xi| < 1$, and if D is completely inside this strip ψ is determined. In the other case, D is divided into a part D' inside the strip and D'' outside. Then the unknown value of ψ_η on D'' can be determined from the integral equation

$$\psi(x, 0, t) = 0 = \iint_{D'} \frac{f(\xi) \, d\xi \, d\tau}{\sqrt{(t - \tau)^2 - (x - \xi)^2}}$$
$$+ \iint_{D''} \frac{\psi_\eta(\xi, 0, \tau) \, d\xi \, d\tau}{\sqrt{(t - \tau)^2 - (x - \xi)^2}} \quad (14)$$

valid for $|x| > 1$.

The problem of the strip is complicated by the fact that D'' may consist of a part of the region $x > 1$ and of $x < -1$, where in both parts ψ_η is unknown. This complication does not exist for the Sommerfeld problem of diffraction by a half-plane, which we solve first in order to check the method.

III. Diffraction by a Half Plane

If the half-plane $x > 0$ is a perfectly reflecting screen, the value of φ_y is given here. The transformed problem then is (Fig. 2)

$$\psi_{xx} + \psi_{yy} - \psi_{tt} = 0 \quad (15)$$

$$\begin{array}{llll} \psi_y = f(x) & y = 0, & x > 0, & t > 0 \\ \psi = 0 & y = 0, & x < 0, & t > 0 \\ \psi = \psi_y = 0 & & & t \leq 0. \end{array}$$
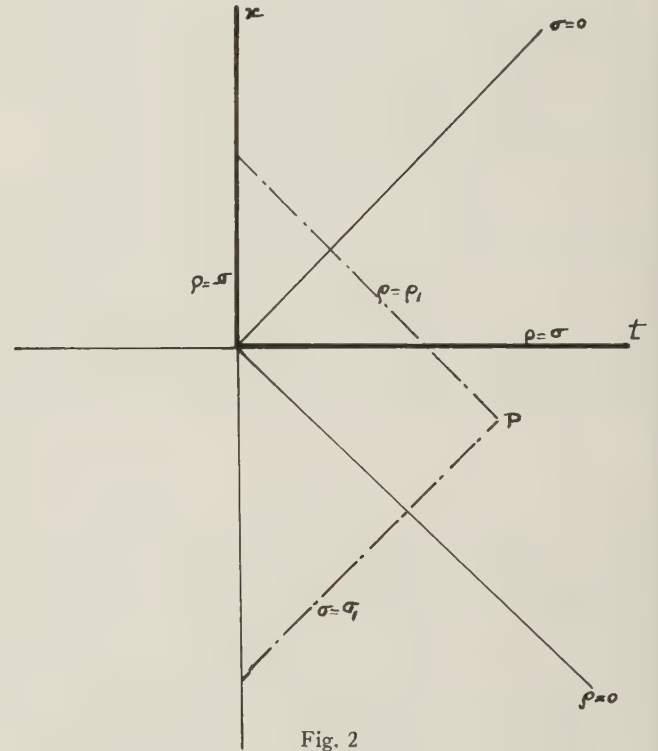
Fig. 2

We first calculate ψ_y for $x < 0$, using the integral equation (14),

$$0 = \iint_{D'} \frac{f(\xi) \, d\xi \, d\tau}{\sqrt{(t - \tau)^2 - (x - \xi)^2}}$$
$$+ \iint_{D''} \frac{\psi_\eta(\xi, \tau) \, d\xi \, d\tau}{\sqrt{(t - \tau)^2 - (x - \xi)^2}}. \quad (16)$$

This equation is rewritten in characteristic coordinates

$$\begin{array}{ll} \rho_1 = t + x, & \rho = \tau + \xi \\ \sigma_1 = t - x, & \sigma = \tau - \xi \end{array}$$

$$0 = \int_0^{\rho_1} \frac{d\rho}{\sqrt{\rho_1 - \rho}} \left[\int_{-\rho}^{+\rho} \frac{f\frac{(\rho - \sigma)}{2}}{\sqrt{\sigma_1 - \sigma}} \, d\sigma \right.$$

$$+ \int_{\rho}^{\sigma_1} \frac{\psi_\eta (\rho, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} \Bigg], \qquad (17)$$

remarking that $\psi_\eta = 0$ in front of the characteristic $\rho = 0$. Since (17) holds identically for all ρ_1 in this range $0 < \rho_1 < \sigma_1$ the integrand vanishes identically here and Evvard[4] obtains the integral equation in one variable σ,

$$\int_{-\rho}^{\rho} \frac{f\left(\dfrac{\rho - \sigma}{2}\right)}{\sqrt{\sigma_1 - \sigma}} \, d\sigma = - \int_{\rho_1}^{\sigma_1} \frac{\psi_\eta (\rho, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}, \qquad (18)$$

which is an Abel integral equation.

For later reference we solve the slightly more general equation

$$\int_{\alpha}^{\sigma 1} \frac{\psi_\eta(\rho, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = - \int_{\beta}^{v} \frac{g(\rho, \sigma) \, d\tau}{\sqrt{\sigma_1 - \sigma}} \qquad (19)$$

by multiplication through $\dfrac{d\sigma_1}{\sqrt{s - \sigma_1}}$ and integration,

$$\int_{\alpha}^{s} \psi_\eta(\rho, \sigma) \, d\sigma \int_{\sigma}^{s} \frac{d\sigma_1}{\sqrt{(s - \sigma_1)(\sigma_1 - \sigma)}}$$

$$= \pi \int_{\alpha}^{s} \psi_\eta(\rho, \sigma) \, d\tau$$

$$= - \int_{\beta}^{v} g(\rho, \sigma) \, d\tau \int_{\alpha}^{s} \frac{d\sigma_1}{\sqrt{(s - \sigma_1) \, (\sigma_1 - \sigma)}}. \qquad (20)$$

Differentiation with respect to s gives

$$w(\rho_1, \sigma_1) = \psi_n(\rho_1, \sigma_1) = \frac{-1}{\pi \sqrt{\sigma_1 - \alpha}}$$

$$\int_{\beta}^{\alpha} \frac{g(\rho_1, \sigma) \sqrt{\alpha - \sigma}}{\sigma_1 - \sigma} \, d\sigma. \qquad (21)$$

In the half-plane case, the solution is

$$w(\rho_1, \sigma_1) = \frac{-1}{\pi \sqrt{\sigma_1 - \rho_1}} \int_{-\rho 1}^{+\rho_1} \frac{f\left(\dfrac{\rho_1 - \sigma}{2}\right) \sqrt{\rho_1 - \sigma}}{\sigma_1 - \sigma} \, d\sigma. \qquad (22)$$

The Fourier-transform of $w(\rho_1, \sigma_1)$ is given by

$$W(x'; k) = \int_{0}^{\infty} e^{ikt} f(x) \, dt = - \frac{1}{ik} f(x) \qquad x > 0 \quad (23)$$

$$W(x, k) = - \frac{1}{\pi \sqrt{2x}} \int_{0}^{\infty} e^{ikt} \int_{-\rho 1}^{+\rho 1} \frac{f\left(\dfrac{\rho_1 - \sigma}{2}\right) \sqrt{\rho_1 - \sigma}}{\sigma_1 - \sigma} \, d\sigma$$

$$x < 0$$

where along the line of integration $x = \text{constant} < 0$.

Putting $\rho_1 - \sigma = 2\xi$, we obtain[5]

$$W(x) = \frac{-1}{\pi i k \sqrt{x}} \int \frac{f(\xi) e^{ikl(\xi - x)} \sqrt{\xi} \, d\xi}{\xi - x}$$

$$= \frac{+1}{\pi \sqrt{x}} \int_{k}^{\infty} d\lambda \int_{0}^{\infty} f(\xi) e^{ikl(\xi - x)} \sqrt{\xi} \, d\xi. \qquad (24)$$

Now in the $(x \cdot y)$ plane the normal derivative φ_y is given on the complete x-axis and we obtain the potential φ from the well-known formula

$$\varphi(x, y; k) = \frac{1}{2k} \int_{-\infty}^{\infty} \varphi_y(\xi) H_o^{(1)}(kr) d\xi \qquad (25)$$

where $r = \sqrt{(x - \xi)^2 + y^2}$, which here gives

$$\varphi(x, y; k) = \frac{1}{2k} \int_{0}^{\infty} f(\xi) H_o^{(1)}(kr) d\xi$$

$$+ \frac{1}{2k\pi} \int_{-\infty}^{0} \frac{d\xi}{\sqrt{\xi}} H_o^{(1)}(kr) \int_{0}^{\infty} \frac{f(\xi) e^{ikl(S - \xi)} \sqrt{\xi} \, ds}{s - \xi}. \qquad (26)$$

Since

$$\sqrt{\xi} \int_{-\infty}^{0} \frac{d_x}{\sqrt{x}} \frac{e^{ikl(\xi - x)}}{\xi - x} H_o^{(1)}\left\{k \sqrt{(x - _x)^2 + y^2}\right\} =$$

$$\frac{2i}{\pi} \int_{\xi + \sqrt{x^2 + y^2}}^{\infty} \frac{e^{ik\lambda} \, d\lambda}{\sqrt{\lambda^2 - (x - \xi)^2 - y^2}}, \qquad (27)$$

this result can be written as

$$\varphi(x, y) = \frac{1}{2k} \int_{0}^{-\infty} f(\xi) \, d\xi \Bigg[H_o^{(1)}(kr)$$

$$+ \frac{2i}{\pi} \int_{\xi + \sqrt{x^2 + y^2}}^{\infty} \frac{e^{ik\gamma} d\lambda}{\sqrt{\lambda^2 - (x - \xi)^2 - y^2}} \Bigg]. \qquad (28)$$

The equivalence to the well-known result of Sommerfeld is demonstrated, if it is shown that the value of φ_y for $x < 0$, $y = 0$ is equal to the value obtained by Sommerfeld.

For a wave incident under an angle ϑ,

$$\varphi^i = e^{ik(x \cos \varphi + y \sin\varphi)} \qquad (29)$$

and

$$f(\xi) = k^2 \sin \vartheta e^{ik\xi \cos \vartheta}. \qquad (30)$$

Then

$$W(x) = \frac{-1}{\pi \sqrt{x}} \int_{k}^{\infty} d\lambda \int_{0}^{\infty} e^{ik\xi \cos \vartheta + i\lambda(\xi - x)} \sqrt{\xi} \, d\xi$$

$$= \frac{e^{\pi/4i} e^{ikx \cos \vartheta}}{\pi \sqrt{x}} \int_{k(1 + \cos \vartheta)}^{\infty} e^{-i\mu x} \mu^{-3/2} \, d\mu \qquad (31)$$

which is easily shown to be equal to $\dfrac{\partial \varphi}{\partial y}$ for Sommerfeld's potential[6].

[4] J. C. Evvard, NACA Rep. 951, Washington; 1950. See also G. N. Ward, *Quart. Jour. Mech. and Appl. Math.*, vol. 2, pp. 136–152, 1949.

[5] Here and in other infinite integrals with respect to λ' we assume along the line of integration $Im\lambda > 0$ in order to secure convergence.
[6] B. B. Baker and E. T. Copson, "The Mathematical Theory of Huygen's Principle," 2nd Ed., p. 142; 1950.

IV. Solution of the Problem for the Slit

The solution to the problem (8), (9) is given by (13), or, in characteristic co-ordinates

$$\psi(\rho_1, \sigma_1) = -\frac{1}{\pi} \iint_D \frac{\psi_\eta \, d\rho \, d\sigma}{\sqrt{(\rho_1 - \rho)(\sigma_1 - \sigma) - y^2}} \quad (32)$$

where D is bounded by $(\rho_1 - \rho)(\sigma_1 - \sigma) = y^2$, $\rho + \sigma = 0$.

Since $\psi_\eta = w(\rho, \sigma) = f\left(\frac{\rho - \sigma}{2}\right)$ is only given for $|\rho - \sigma| < 2$, the value of w in the two side regions $|\rho - \sigma| > 2$, must be determined from the integral equations

$$0 = \psi(\rho_1, \sigma_1) = -\frac{1}{\pi} \iint_{D'} \frac{f \, d\rho \, d\sigma}{\sqrt{(\rho_1 - \rho)(\sigma_1 - \sigma)}}$$
$$-\frac{1}{\pi} \iint_{D''} \frac{\psi_\eta(\rho, \sigma) \, d\rho \, d\sigma}{\sqrt{(\rho_1 - \rho)(\sigma_1 - \sigma)}}, \quad (33)$$

where D' lies inside the strip and D'' outside.

For $\rho < -1 \, \psi_y = \omega$ is identically zero, for $-1 < \rho < 1$ w can be calculated directly from the known values on the strip, but for $1 < \rho < 3$ the calculation involves the values for $-1 < \sigma < 1$, which also are directly calculated (Fig. 3). In the next region $3 < \rho < 5$ the
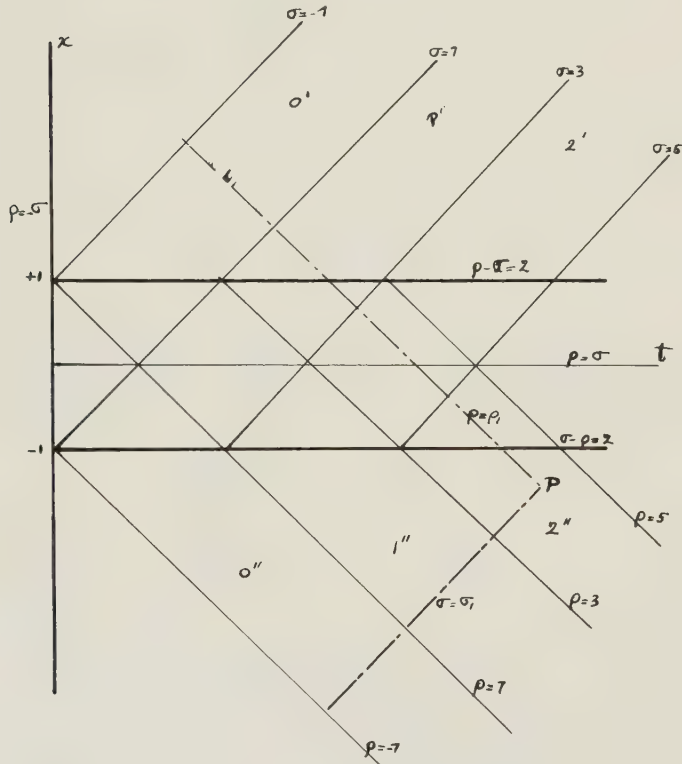
Fig. 3

values follow from the known values in the region $1 < \sigma < 3$ and so it is seen, that dividing the two side regions into sub-regions the values in each subregion can be expressed by previously calculated values. We now derive the recursion formula for this procedure.

We denote the regions in the following way:

$\rho - \sigma > 2$		$\sigma - \rho > 2$	
$0'$	$-1 < \sigma < 1$	$0''$	$-1 < \rho < 1$
$1'$	$1 < \sigma < 3$	$1''$	$1 < \rho < 3$
$2'$	$3 < \sigma < 5$	$2''$	$3 < \rho < 5$
n'	$2n - 1 < \sigma < 2n + 1$	n''	$2n - 1 < \rho < 2n + 1$

The integral equation in region n'' for $w''(\rho_1, \sigma_1)$, where $2n - 1 < \rho_1 < 2n + 1$, $\sigma_1 > \rho_1 + 2$ is

$$0 = \psi(\rho_1, \sigma_1) = -\frac{1}{\pi} \int_{-1}^{\rho_1} \frac{d\rho}{\sqrt{\rho_1 - \rho}} \left\{ \int_{-1}^{\rho-2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{(\sigma, -\sigma)}} \right.$$
$$\left. + \int_{\rho-2}^{\rho+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} + \int_{\rho+2}^{\sigma_1} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} \right. \quad (34)$$

Again this integral equation is reduced to an integral equation in one variable

$$\int_{-1}^{\rho_1-2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} + \int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}$$
$$+ \int_{\rho_1+2}^{\sigma_1} \frac{\omega(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = 0. \quad (35)$$

We now build up the functions $w'(\rho_1, \sigma_1)$ and $w''(\rho_1, \sigma_1)$ in the successive regions as sums of solutions of integral equations, which become more complicated at each step. We define for all values $\rho_1 > -1$ and $\sigma_1 > 2 + \rho_1$ $w_o''(\rho_1, \sigma_1)$ as the solution of the equation

$$\int_{\rho_1+2}^{\sigma_1} \frac{w_o''(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{\rho_1+2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} \quad (36)$$

and $w_o'(\rho_1, \sigma_1) = 0$ for $\rho_1 < -1$. Here $w = 0$ for $\sigma < -\rho_1$. w_o' is defined in a similar way, interchanging ρ and σ. w_1' is the correction to w_o' for values of ρ_1 with $1 < \rho_1 < 3$ and $w_1' = 0$ for $\rho_1 < 1$. Hence

$$\int_{\rho_1+2}^{\sigma_1} \frac{(w_o'' + w_1'') \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}$$
$$- \int_{-1}^{\rho_1+2} \frac{w_o'(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}. \quad (37)$$

Subtraction of (37) and (36) gives

$$\int_{\rho_1+2}^{\sigma_1} \frac{w_1''(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{-1}^{\rho_1-2} \frac{w_o'(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}. \quad (38)$$

We define $w_1''(\rho_1, \sigma_1)$ as the solution of this equation for all values of $\rho_1 > 1$ and $w'' = 0$ for $\rho_1 < 1$. Similarly we put, for $2n - 3 < \rho_1 < 2n - 1$

$$w''(\rho_1, \sigma_1) = \sum_{k=0}^{n-1} w_k'',$$

where $w_{n-1}'' = 0$ for $F \rho_1 < 2n - 3$ and w_{n-1}'' is the solution of

$$\int_{\rho_1+2}^{\sigma_1} \frac{\left(\sum_{k=0}^{n-1} w_k''\right) d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}$$

$$-\int_{-1}^{2n-5} \frac{\left(\sum_{k=0}^{n-2} w_k'\right) d\sigma}{\sqrt{\sigma_1 - \sigma}} - \int_{2n-5}^{\rho_1-2} \frac{\left(\sum_{k=0}^{n-2} w_k'\right) d\sigma}{\sqrt{\sigma_1 - \sigma}} \quad (39)$$

for all values of $\rho_1 > 2n - 3$.

Also, $w_n''(\rho_1, \sigma_1) = 0$ for $\rho_1 < 2n - 1$ and for all values $\rho_1 > 2n - 1$ it is the solution of

$$\int_{\rho_1+2}^{\sigma_1} \frac{\left(\sum_{k=0}^{n} w_k''\right) d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}$$

$$-\int_{-1}^{2n-3} \frac{\left(\sum_{k=0}^{n-1} w_k'\right) d\sigma}{\sqrt{\sigma_1 - \sigma}} - \int_{2n-3}^{\rho_1-2} \frac{\left(\sum_{k=0}^{n-1} w_k'\right) d\sigma}{\sqrt{\sigma_1 - \sigma}}. \quad (40)$$

Subtraction of these two equations gives for $\rho_1 > 2n - 1$,

$$\int_{\rho_1+2}^{\sigma_1} \frac{w_n''(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}} = -\int_{2n-3}^{\rho_1-2} \frac{w_{n-1}'(\rho_1, \sigma) \, d\sigma}{\sqrt{\sigma_1 - \sigma}}. \quad (41)$$

Considering this as an integral equation for $w_n''(\rho_1, \sigma)$ it can be solved by (21),

$$\omega_n''(\rho_1, \sigma_1) = -\frac{1}{\pi\sqrt{\sigma_1 - \rho_1 - 2}}$$

$$\int_{2n-3}^{\rho_1-2} \frac{w_{n-1}'(\rho_1, \sigma) \sqrt{\rho_1 + 2 - \sigma} \, d\sigma}{\sigma_1 - \sigma} \quad (42)$$

Interchanging of ρ and σ yields the recursion formula for the other side region,

$$w_n'(\rho_1, \sigma_1) = -\frac{1}{\pi\sqrt{\rho_1 - \sigma_1 - 2}}$$

$$\int_{2n-3}^{\sigma_1-2} \frac{w_{n-1}''(\rho, \sigma_1) \sqrt{\sigma_1 + 2 - \rho}}{\rho_1 - \rho} \, d\sigma. \quad (43)$$

The first terms are given by

$$w_o'(\rho_1, \sigma_1) = -\frac{1}{\pi\sqrt{\rho_1 - \sigma_1 - 2}}$$

$$\int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho, \sigma_1) \sqrt{\sigma_1 + 2 - \rho}}{\rho_1 - \rho} \, d\rho \quad (44)$$

$$w_o''(\rho_1, \sigma_1) = -\frac{1}{\pi\sqrt{\sigma_1 - \rho_1 - 2}}$$

$$\int_{\rho_1-2}^{\rho_1+2} \frac{w(\rho_1, \sigma) \sqrt{\rho_1 + 2 - \sigma}}{\sigma_1 - \sigma} \, d\sigma \quad (45)$$

with $w(\rho, \sigma) = 0$ for $\rho + \sigma < 0$. Then

$$\psi_v = w'(\rho, \sigma) = \sum_{n=0}^{\infty} \omega_n'(\rho, \sigma) \quad \rho - \sigma > 2, \quad \sigma > -1 \quad (46)$$

$$\psi_v = w''(\rho, \sigma) = \sum_{n=0}^{\infty} \omega_n''(\rho, \sigma) \quad \sigma - \rho > 2, \quad \rho > -1$$

The value of φ_y, the normal derivative of the potential of the original diffraction problem is found by the Laplace transform

$$W_n'(x) = \int_0^{\infty} e^{ikt} w_n'(\rho, \sigma) \, dt =$$

$$\int_{2n-1}^{\infty} e^{ik(x+\sigma)} w_n'(\sigma + 2x, \sigma) \, d\sigma, \quad (47)$$

where along the line of integration $\rho - \sigma = 2x$.

Then

$$W'(x) = \sum_{n=0}^{\infty} W_n'(x) \quad x > 1, y = 0,$$

$$W''(x) = \sum_{n=0}^{\infty} W_n''(x) \quad x < -1, y = 0.$$

The recursion formula for $W(x)$ is found by transformation

$$W_n'(x) = \int_{2n-1}^{\infty} e^{ik(n+\omega)} w_n'(\sigma + 2x, \sigma) \, d\sigma =$$

$$-\frac{e^{ikx}}{\pi\sqrt{2(x-1)}} \int_{2n-1}^{\infty} e^{ik\sigma} \, d\sigma$$

$$\int_{2n-3}^{\sigma-2} \frac{\omega_{n-1}''(\rho, \sigma) \sqrt{\sigma + 2 - \rho}}{\sigma + 2x - \rho} \, d\rho. \quad (48)$$

Putting $\rho - \sigma = 2\xi$, and interchanging the order of integration,

$$W_n'(x) = -\frac{1}{\pi\sqrt{x-1}}$$

$$\int_{-1}^{-\infty} \frac{e^{ik(x-\xi)} W_{n-1}''(\xi)\sqrt{1-\xi}}{x-\xi} \, d\xi =$$

$$\frac{i}{\pi\sqrt{x-1}} \int_k^{\infty} d\lambda \int_{-1}^{-\infty} e^{i\lambda(x-\xi)} W_{n-1}''(\xi; k) \sqrt{1-\xi} \, d\xi, \quad (49)$$

and similarly

$$W_n''(x) = -\frac{1}{\pi\sqrt{-x-1}} \int_1^{\infty} \frac{e^{ik(\xi-x)} W_{n-1}'(\xi)\sqrt{1+\xi}}{\xi - x} \, d\xi$$

$$= \frac{i}{\pi\sqrt{-x-1}} \int_k^{\infty} d\lambda \int_1^{\infty} e^{i\lambda(\xi-x)} W_{n-1}'(\xi, k) \sqrt{1+\xi} \, d\xi. \quad (50)$$

The transforms $W_o'(x; k)$ and $W_o''(x; k)$ are given by

$$W_o'(x, k) = \frac{-1}{\pi k\sqrt{x-1}} \int_k^{\infty} d\lambda \int_{-1}^{+1} e^{i\lambda(x-\xi)} w(\xi)\sqrt{1-\xi} \, d\xi$$

$$W_o''(x, k) = \frac{-1}{\pi k \sqrt{-x-1}} \int_k^\infty d\lambda \tag{51}$$

$$\int_k^\infty d\lambda \int_{-1}^{+1} e^{i\lambda(\xi-x)} w(\xi) \sqrt{1+\xi} \, d\xi. \tag{52}$$

The value of φ in an arbitrary point (x, y) is

$$\varphi(x, y) = \frac{1}{2k} \int_{-1}^{+1} f(\xi) H_o^{(1)}(kr) \, d\xi + \tag{53}$$

$$\frac{1}{2i} \left[\int_{-\infty}^{-1} W''(\xi) H_o^{(1)}(kr) \, d\xi + \int_1^\infty W'(\xi) H_o^{(1)}(kr) \, d\xi \right],$$

where

$$r = \sqrt{(x-\xi)^2 + y^2}.$$

V. The Transmission Cross Section

The transmission cross section for a plane wave incident at an angle ϑ can be calculated by a formula of Levine and Schwinger[7]

$$\sigma(\vartheta) = -\frac{2\pi}{k} \, Im\{A(\vartheta)\}, \tag{54}$$

where $A(\vartheta)$ is the amplitude of the scattered wave at infinity in the direction of the normal of the incoming wave, which is defined by

$$\varphi(\vartheta) \sim \sqrt{\frac{2\pi}{kr_o}} \, e^{i(kr_o + \pi/4)} A(\vartheta), \tag{55}$$

$r_o = \sqrt{x^2 + y^2}$ being the distance to the center of the aperture. Using the asymptotic expression for the Hankel function

$$H_o^{(1)}(kr) \sim \sqrt{\frac{2}{\pi kr}} e^{i(kr - \pi/4)} \sim \sqrt{\frac{2}{\pi kr_o}} e^{ikr_o - i\pi/4} e^{ik\xi \cos\vartheta}, \tag{56}$$

(53) yields

$$A(\vartheta) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\partial \varphi(\xi)}{\partial \eta} e^{ik\xi \cos\vartheta} \, d\xi =$$

$$\frac{-i}{2k\pi} \int_{-1}^{+1} f(\xi) e^{ik\cos\vartheta} \, d\xi - \frac{1}{2\pi} \left[\int_{-\infty}^{-1} W''(\xi) e^{ik\xi \cos\vartheta} \, d\xi + \int_1^\infty W'(\xi) e^{ik\xi \cos\vartheta} \, d\xi \right] \tag{57}$$

The first term gives the Kirchof approximation, the others are the correction terms.

Since the calculations turn out to be rather laborious we perform here these calculations only for the case of normal incidence. Then, from (30),

$$w(\xi) = f(\xi) = k^2 \tag{58}$$

and

_____

[7] C. J. Bouwkamp, "Diffraction Theory," *Rep. Prog. Phys.*

$$A(o) = -\frac{2ik}{2\pi} - \frac{1}{2\pi} \left[\int_{-\infty}^{-1} W''(\xi) \, d\xi + \int_1^\infty W'(\xi) \, d\xi \right]. \tag{59}$$

The first term gives

$$\sigma^{(o)} = 2. \tag{60}$$

The functions $W_n'(x)$ are calculated from (49)–(52);

$$W_o'(x) = \frac{-k}{\pi \sqrt{x-1}} \int_k^\infty d\lambda \int_{-1}^{+1} e^{i\lambda(x-\xi)} \sqrt{1-\xi} \, d\xi. \tag{61}$$

Introducing the function

$$F_n(u) = \int_u^\infty e^{iv} v^{n-\frac{1}{2}} \, dv, \tag{62}$$

which is related to Fresnel's integrals, we write

$$W_o'(x) = \frac{-k}{\pi \sqrt{x-1}} \int_k^\infty e^{i\lambda(x-1)} \lambda^{-3/2} d\lambda \{F_1(o) - F_1(2\lambda)\}. \tag{63}$$

$W_1''(x)$ is now calculated, introducing the function

$$H(\lambda + \mu) = \int_1^\infty e^{i(\lambda+\mu)\xi} \frac{1+\xi}{\sqrt{\xi^2-1}} \, d\xi$$

$$= \frac{\pi i}{2} \left\{ H_o^{(1)}(\lambda + \mu) + i H_1^{(1)}(\lambda + \mu) \right\}, \tag{64}$$

which gives

$$W_1''(x) = \frac{-ik}{\pi^2 \sqrt{-1-x}} \left\{ F_1(o) \cdot \int_k^\infty e^{-i\lambda x} \, d\lambda \right.$$

$$\int_k^\infty e^{-i\mu} \mu^{-3/2} H(\lambda + \mu) \, d\mu -$$

$$\left. - \int_k^\infty e^{-i\lambda x} \, d\lambda \int_k^\infty e^{-i\mu} \mu^{-3/2} H(\lambda + \mu) F_1(2\mu) \, d\mu \right\}. \tag{65}$$

The higher terms are found by successive application of the recursion formula

$$W_{2n}'(x) = \frac{-1}{\pi^{2n} \sqrt{x-1}} \int_k^\infty e^{i\lambda x} \, d\lambda \; .$$

$$\int_k^\infty d\mu_1 H(\lambda + \mu_1) \cdots \int_k^\infty dv_n H(\mu_{2n-2} + v)$$

$$\int_1^\infty e^{iv\xi} W_o'(\xi) \sqrt{1+\xi} \, d\xi \tag{66}$$

$$W_{2n+1}''(x) = \frac{-1}{\pi^{2n} \sqrt{-x-1}} \int_k^\infty e^{-i\lambda x} \, d\lambda$$

$$\int_k^\infty d\mu_1 H(\lambda + \mu_1) \cdots \int_k^\infty dv H(\mu_{2n-1} + v)$$

$$\int_1^\infty e^{iv\xi} W_1'(\xi) \sqrt{1+\xi} \, d\xi. \tag{67}$$

From these results asymptotic formulae for large values of k can be derived, using the expansions

$$\dot{F}_n(u) \sim ie^{iu}u^{n-1/2}\left[1 + \frac{i(n-\frac{1}{2})}{u} - \right.$$

$$\left. \frac{(n-\frac{1}{2})(n-\frac{3}{2})}{u^2} - i\frac{(n-\frac{1}{2})(n-\frac{3}{2})(n-\frac{5}{2})}{u^3} \cdots \right]. \quad (68)$$

$$H(p) = e^{i(\rho+\pi/4)}\sqrt{\frac{2\pi}{\rho}}\left[1 + \frac{i}{8p} + \right.$$

$$\left. \frac{3}{32p^2} - \frac{15i}{12\vartheta p^3} + \cdots \right]. \quad (69)$$

Since in (66) and (67) both λ and μ are $> k$, it is clear from this expansion that each term is of order $k^{1/2}$ compared to the preceding; hence the series is an asymptotic series. We now calculate the contributions of the successive terms to the transmission cross section. Owing to the symmetry

$$T_n = T_n' = T_n'' = \int_1^\infty W_n'(x)\,dx.$$

Then

$$T_o' = \frac{-k}{\pi}\int_1^\infty \frac{dx}{\sqrt{x-1}}\int_k^\infty e^{i\lambda(x-1)}\lambda^{-3/2}d\lambda\{F_1(o) - F_1(2\lambda)\} =$$

$$- \frac{1}{2} + \frac{e^{\pi/4i}}{\sqrt{\mu}}\{F_1(2k) - 2kF_o(2k)\} \quad (70)$$

$$T_1'' = \frac{ie^{\pi/4i}}{\pi^{3/2}}k\cdot\int_k^\infty d\lambda\int_k^\infty d\mu\cdot$$

$$e^{i(\lambda-\mu)}H(\lambda + \mu)\mu^{-3/2}\lambda^{-1/2}\left[\frac{e^{3/4\pi i}}{2}\sqrt{\pi} - F_1(2\mu)\right] \quad (71)$$

$$T_2' = \frac{-e^{\pi/4i}}{\pi^{5/2}}k\int_k^\infty d\lambda\int_k^\infty d\mu e^{i(\lambda-\mu)}H(\lambda + \mu)\cdot$$

$$\int_k^\infty H(\mu + v)v^{-3/2}\lambda^{-1/2}\left[\frac{e^{3/4\pi i}}{2}\sqrt{\pi} - F_1(2v)\right] \quad (72)$$

Inserting the asymptotic series, (68) and (69) give, after some lengthy calculations, the amplitude

$$A(o) = -\frac{ik}{\pi} + \frac{1}{2\pi} + \frac{e^{i(2k-\pi/4)}}{2\pi\sqrt{\pi k}} - \frac{ie^{4ik}}{4\pi^2 k} + 0(k^{-3/2}); \quad (73)$$

from which the cross section is derived

$$\sigma(o) = 2 - \frac{2\sin(2k + \pi/4)}{\sqrt{\pi k^3}} + \frac{\cos 4k}{2\pi k^2} + 0(k^{-5/2}). \quad (74)$$

The results of these formulas are compared with the results, obtained by Skavlem[8],

TABLE I

k	(exact)	(asymptotic)
1	0.5454	0.6992
2	1.1843	1.1932
3	0.9720	0.9627
4	0.9424	0.9484
5	1.0499	1.0519
6	0.9956	0.9935
7	0.9717	0.9726
8	1.0233	1.0240
9	1.0020	1.0011
10	0.9822	0.9823

[8] S. Skavlem, *Arch. Math. Naturv.*, vol., 51, pp, 61-80; 1951.

The Modeling of Physical Systems

R. K. RITT†

I

TO DISCOVER the properties of a physical system it is necessary, in general, to compute or measure certain functions. These functions will depend upon space and time variables, upon certain physical parameters, and upon initial or boundary conditions; computing them is usually a matter of solving differential, integral, or differentio-integral equations. The alternative to computing them is to measure them. Here the difficulties are formidable, and resort is often made to modeling. The idea behind modeling is simple: change the values of the physical parameters, change the initial or boundary conditions, change the space and time variables, in such a manner that an experiment becomes feasible; then measure the crucial functions and from these measurements determine the functions of the original system. It is not true, however, that these changes can be performed arbitrarily; certain restrictions automatically enter, and are referred to as conditions of similitude. It is these similitude conditions, for a certain class of systems, which we propose to investigate.

II

To illustrate the central idea, let us examine a classical situation.[1] Suppose the system at hand is described by a linear differential equation

$$\frac{dv}{dt} = Av \qquad (1)$$

where v is an n-vector, and A is a matrix which is independent of t. Let

$$\frac{du}{ds} = Bu, \qquad (2)$$

where u is an n-vector, and B is a matrix which is independent of s. Suppose, further, that $s = \lambda t$ and let T be a constant matrix. If the experimental measurement of u is feasible, let us determine whether it is possible to compute v by setting $v = Tu$. We secure:

$$\lambda\, TBu = \lambda \frac{d}{ds}(Tu) = \lambda \frac{dv}{ds} = \frac{dv}{dt} = Av = ATu \qquad (3)$$

so that

$$\lambda\, TB = AT. \qquad (3')$$

This last equation imposes relationships on the elements of A, B, and T; these relationships are the equations of similitude. Now, it should be remarked that in this illustration no discussion of initial conditions occurred. This is accounted for by the *a priori* choice of the linear relationship between u and v. In fact, if the

above discussion is to be meaningful, u_o, the initial value of u, must be the same as Tv_o, where v_o is the initial value of v.

III

Now suppose, referring to the same systems, (1) and (2), we defer our choice of the relationship between u and v. Instead, we choose initial conditions u_o, v_o arbitrarily, and write the equations

$$v = \exp(At)\, v_o \qquad (1')$$
$$u = \exp(Bs)\, u_o \qquad (2')$$

where $\exp(At)$ is the matrix operator $\sum_{n=o}^{\infty} \frac{t^n A^n}{n!}$.

Eqs. (1') and (2') are the solutions of (1) and (2) and describe trajectories in the n-dimensional space. Given any point on the trajectory (2') corresponding to s, we associate the point on the trajectory (1') corresponding to $t = s/\lambda$. The question we ask is: does this association define a function from (2') to (1')? If the answer is affirmative, we shall call this function the *modeling function*, and any measurement of u can be used to compute a value of v. In general, the computation of this function is not to be expected to be less difficult than the solution of the original differential equations, but it is possible to give examples where, formally at least, it is a *different* problem, and a program of discovering when it is an easier problem is projected. There is one case in which the computation of the modeling function is trivial, but which has implications in electromagnetic and acoustical modeling which, as far as we know, are novel. Suppose that $n = 1$, so that A, B, u_o, v_o are numbers rather than matrices and vectors. Then the function we are seeking is

$$v = v_o \left(\frac{u}{u_o}\right)^{A/\lambda B}. \qquad (4)$$

We shall see, below, that (4) has its counterpart in the modeling of elliptic systems.

IV

We shall now concern ourselves with the *existence* of the modeling function. In the case at hand the necessary and sufficient condition that the modeling function exist is

$$\exp(Bs)\, u_o = u_o \rightarrow \exp(As/\lambda)\, v_o = v_o. \qquad (5)$$

Now, let σ be the first nonzero value of s for which the left side of (5) is satisfied. If $\sigma = \infty$, the modeling function exists. If σ is finite $\exp(B\sigma)\, u_o = u_o$, which means that u_o can be written

$$u_o = \sum_{j=1}^{p} c_j u_j \; ; \; Bu_j = \xi_j u_j \; ; \; \exp(\xi_j \sigma) = 1, \qquad (6)$$

where the c_j are constants.

---

† Univ. of Michigan, Ann Arbor, Mich.
[1] The detailed proofs of the statements made in Sections II, III and IV will appear elsewhere.

Then (5) will be satisfied if and only if

$$v_o = \sum_{j=1}^{r} d_j v_j \; ; \; A v_j = \eta_j v_j \; ; \; \exp{(\eta_j \sigma / \lambda)} = 1 \qquad (6')$$

where the d_j are constants.

From this, several conclusions can be drawn. First, the modeling function will certainly exist, regardless of the choice of u_o and v_o, if B has no purely imaginary characteristic values. If B has purely imaginary characteristic values, if $\xi_1, \xi_2, \ldots, \xi_p$ are such characteristic values, which are commensurate, and if u_o is a linear combination of the corresponding characteristic vectors, then σ will be the smallest number such that $\sigma \xi_j$ is an integral multiple of $2\pi i$ for $j = 1, \ldots, p$. If the modeling function is to exist, then A must have characteristic values $\eta_1, \eta_2, \ldots, \eta_r$ which are such that $\eta_j \sigma / \lambda$ is an integral multiple of $2\pi i$, $j = 1, \ldots, r$, and v_o must be a linear combination of the corresponding characteristic vectors. For example, if

$$A = \begin{pmatrix} 0 & 1 \\ -k^2 & 0 \end{pmatrix}; \quad B = \begin{pmatrix} 0 & 1 \\ -l^2 & 0 \end{pmatrix},$$

then u_o is of necessity a combination of the characteristic vectors corresponding to $\pm li$, so $\sigma = 2\pi / l$; v_o is a combination of the characteristic vectors corresponding to $\pm ki$; therefore, k, l, and λ must satisfy the relation $k = \lambda n l$, n an integer. In this case, if u and v have components $u^{(1)}$, $u^{(2)}$, and $v^{(1)}$, $v^{(2)}$ respectively, it is easy to verify that the modeling function can be written in terms of polynomials:

$$v^{(1)} = v_o^{(1)} \cos{[n \cos^{-1} P]} + \frac{v_o^{(2)}}{k} \sin{[n \sin^{-1} Q]}$$

$$v^{(2)} = v_o^{(2)} \cos{[n \cos^{-1} P]} - k v_o^{(1)} \sin{[n \sin^{-1} Q]} ,$$

where P and Q are functions of l, $u^{(1)}$, $u^{(2)}$, $u_c^{(1)}$, $u_o^{(2)}$. The case $n = 1$ expresses the classical dependence of the period of a harmonic oscillator on the spring constant.

The central fact which emerges from this discussion is that in order to decide whether the system (2) can be used to model system (1), it is sufficient to examine the characteristic values of the matrices A and B, and the composition of the initial vectors v_o and u_o in terms of the characteristic vectors of these matrices.

V

We now turn our attention to a wider class of systems to which most of the above remarks continue to apply.[2] Let V be a function of the variable t, and other variables which we will indicate by Q, so we can let $V = V(t; Q)$. Suppose it is required that V satisfy a differential equation of the form:

$$L_1(V) + L_2(V) = 0 \qquad (7)$$

where L_1 and L_2 are linear differential operators, L_1 involving differentiation with respect to t, L_2 involving differentiation with respect to the remaining variables.

[2] The general results about the Huygens-Hadamard principle and semigroups which are referred to in this and the next section can be obtained from E. Hille, *Functional Analysis and Semi-groups*, American Math. Soc. Col. Pub. XXXI, ch. XX; 1948.

Further, suppose that the coefficients in L_1 are constant. Suppose that to well-set the "Cauchy" or "Dirichlet" problem, it is sufficient to specify

$$V, \frac{\partial V}{\partial t}, \ldots, \frac{\partial^n V}{\partial t^n}$$

as $t \to 0$. Then the Huygens-Hadamard principle is applicable.

This can be interpreted in the following fashion: Let S be a linear vector space consisting of the n-fold product of a space of functions of Q which are sufficiently differentiable that L_2 can be applied to them, and that this space is given a norm of any of the standard forms. Then if v_o is the element of S whose components are the "initial" conditions, there exists a one parameter semigroup T_t of bounded operators on S such that the solution of (7) is given as the first component of the vector

$$v(t, Q) = T_t v_o(Q) . \qquad (8)$$

It is assumed that some suitable norm has been defined for S; it is then known that (8) can be expressed as

$$v(t, Q) = \exp{(At)} v_o(Q) , \qquad (8')$$

where A is, in general, an unbounded operator and $\exp{(tA)}$ must be interpreted differently than before. What is essential to the present discussion is that, when the correct interpretation of $\exp{(At)}$ is made, the statements of sections III and IV continue to apply. This is because the relationships between the spectral theory of the operators A and $\exp{(At)}$ carry over from the finite dimensional to the infinite dimensional case, as long as we confine our attention to the point spectrum.[3] Thus, to study the possibility of using one such system to model another, we need only study the infinitesimal generators of the corresponding semigroups.

VI

If (7) is second order, and either hyperbolic or parabolic, we can find A directly from inspection of (7). If, however, (7) is elliptic, special arguments must be employed. However, since in the elliptic case, a single initial condition may set the problem, a situation analogous to that which produced (4) may arise, so that in some cases, the modeling function is easy to compute.

Consider the equation:

$$\frac{\partial^2 v}{\partial t^2} + k^2 v + \nabla^2 v = 0 , \quad \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} . \qquad (9)$$

If v is required to satisfy a radiation condition, and if $v_o = v_o(x, y)$, the solution of (9) is given

$$v(t, x, y) = -\frac{1}{2\pi} \frac{\partial}{\partial t} \iint v_o(x + x', y + y') G_t(x', y') \, dx' dy' \qquad (10)$$

where $G_t(x, y) = (t^2 + x^2 + y^2)^{-1/2} \exp{[-ik(t^2 + x^2 + y^2)^{1/2}]}$. Changing to polar co-ordinates, integrating by parts, and differentiating gives:

$$v(t, x, y) = \exp{(-ikt)} v_o + t \int_0^\infty G_t(r) \frac{\partial M}{\partial r} \, dr, \qquad (10')$$

[3] *Ibid.*, theorem 15.6.2.

where $M = \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} v_o\,(x+r\cos\theta,\, y+r\sin\theta)\,d\theta$. To find the operator A, we compute

$$\lim_{t\to 0} \frac{\partial v}{\partial t}\,(t,x,y).$$

This gives

$$Av_o = -\,i\,k\,v_o + \int_0^\infty \frac{\exp(-ikr)}{r}\,\frac{\partial M}{\partial r}\,dr\,. \qquad (11)$$

Now,

$$\frac{\partial M}{\partial r} = \frac{1}{2\pi r}\int_0^{2\pi}\big[\nabla v_o\,(x+r\cos\theta,\,\mathrm{y}+r\sin\theta)\big]\cdot \mathrm{n}r\,d\theta$$

$$= \frac{1}{2\pi r}\iint_{\mathrm{Gr}} \nabla^2 v_o\,d\,G$$

where G_r is a circle of radius r, center at (x,y). Thus

$$Av_o = -\,i\,k\,v_o + \int_0^\infty \frac{\exp(-ikr)}{2\pi r^2}\left(\iint_{\mathrm{Gr}} \nabla^2 v_o\, dG\right)dr. \quad (11')$$

From (11') it is immediately evident that $-ik$ is a characteristic value of A, and any harmonic function will be a characteristic vector corresponding to $-ik$. This gives the result that if

$$\frac{\partial^2 u}{\partial s^2} + \nabla^2 u + l^2 u = 0 \qquad (9')$$

is to be used as a model for (9), and if u_o and v_o are harmonic functions, then $k = \lambda n l$, as in the case of the harmonic oscillator. But we can say even more: it is a simple computation to produce the modeling function

$$v(t,x,y) = v_o(x,y)\left[\frac{u(t,x,y)}{u_o(x,y)}\right]^n. \qquad (3')$$

To find the rest of the spectrum of A, we secure, formally, from (8') and (9),

$$A^2 = -\,(\nabla^2 + k^2)\,. \qquad (12)$$

This equation could be verified directly (and laboriously) from (11'). But since the point spectrum of ∇^2 (in the space of bounded, twice differentiable functions) is the negative real axis, the point spectrum of A^2 is the real axis to the right of $-k^2$. But this means, from the spectral mapping theorem,[4] that every purely imaginary characteristic value of A is equal to or less than k in absolute value, and if p is a number $0 \le p \le k$, then at least one of the two numbers $\pm pi$ is in the point spectrum of A. The corresponding characteristic vectors are annulled by the operator

$$\nabla^2 + (k^2 - p^2)\,. \qquad (13)$$

These remarks make it possible to answer the question of the existence of a modeling function for the systems (9) and (9'), for any pair of initial conditions u_o and v_o.[5]

VII

If the scalar wave equation is written in spherical co-ordinates, and the initial data is given on a sphere, it is no longer possible to find a parameter, t, such that the surface of the sphere is given as $t = 0$, and the wave equation can be written as an equation with constant coefficients with respect to differentiation with respect to t. The single exception occurs when $k = 0$, in which case $t = \log(r/a)$, where a is the radius of the sphere. Thus the Huygens-Hadamard principle does not apply, and the analog of the argument in section VI does not, apparently, exist. This seems to the author a peculiar circumstance, and deserves further investigation.

---

[4] *Ibid.*, pp. 312–316.

[5] It is of some interest to note that here is an instance of a solution of the three-dimensional wave equation which satisfies, on the boundary, the two-dimensional wave equation. This phenomenon is referred to by B. Van der Pol, "On discontinuous electromagnetic waves and the occurrence of a surface wave," paper 2.15 in this issue.

On the Diffraction Field Near a Plane-Screen Corner

W. BRAUNBEK†

Summary—It is shown that the diffraction field of an incident plane scalar wave in the vicinity of a plane-screen corner of arbitrary angle can be found approximately by solving Laplace's equation. An approximate solution, which satisfies the boundary conditions exactly, is presented as a simple closed expression by generalizing the known solution of the half-plane problem. A special corner condition, in addition to Meixner's edge condition, is not necessary.

INTRODUCTION

CONSIDER a plane scalar wave $u_1 = \exp(-ikz)$, propagated in the $+z$ direction (the time factor being $\exp(i\omega t)$) normally incident upon a perfectly-reflecting, plane screen of zero thickness, which lies in the x,y plane and has the boundary condition $u = 0$ on both sides. The screen has the form of an infinitely extended sector with the angle $2\vartheta_o$ (Fig. 1) and is symmetrical about the x axis. A point P of space is defined by the spherical co-ordinates r, ϑ, φ, the x axis being the zero direction for ϑ and the $x,+y$ half-plane being the zero half-plane for φ. Instead of ϑ we use as a new variable $p = \cos \vartheta$ and define $p_o = \cos \vartheta_o$.
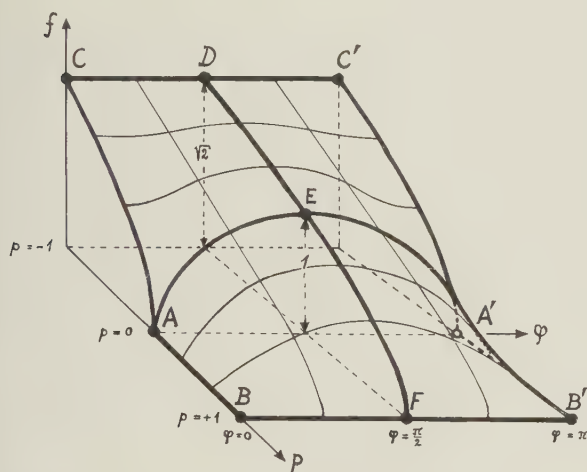
Fig. 1—The sector screen with the symbols used. Left top corner a screen with a re-entering corner.

The problem is to find the diffraction field $u(r,p,\varphi)$ uniquely defined by the conditions stated above, especially in the immediate vicinity of the corner 0 of the screen, for $kr \ll 1$, at a distance from 0 small compared with the wavelength.

The result is interesting not only for the idealized screen here introduced, but generally for the behavior of a scalar wave at a salient or re-entering corner of a plane-screen aperture bordered by an arbitrary broken line—as long as one confines the solution to a region around the corner which is small compared with the

wavelength and small compared with the distance of the next corner and with the radii of curvature of the border lines.

We put the total field $(u_1 + u)$. Then the additional field u due to the screen has to satisfy the following conditions:

$$\Delta u + k^2 u = 0 \qquad \text{everywhere.} \qquad (1)$$

u symmetrical about the x,y plane
$$\text{or: } u(r,p,-\varphi) = u(r,p,\varphi) \; ; \; (0 \leqslant \varphi \leqslant \pi). \qquad (2)$$

u symmetrical about the x,z plane
$$\text{or: } u(r,p, \pi-\varphi) = u(r,p,\varphi). \qquad (3)$$

$$u = 0 \text{ for } \varphi = 0 \text{ or } \pi \text{ and } \vartheta \leqslant \vartheta_o \; (p \geqslant p_o). \qquad (4)$$

Moreover u has to obey Sommerfeld's radiation condition[1] at infinity and Meixner's edge condition[2] at the edges of the screen. Whether an additional "corner condition" at the corner 0 has to be introduced will be investigated later.

By separation of the variable r we obtain as a solution of (1):

$$u = n \sum f_n(p,\varphi) \, r^{-1/2} \, Z_{\alpha_n+\frac{1}{2}}(kr)$$
$$= n \sum f_n(p,\varphi) \sum_{s=0}^{s=\infty} C_s \, r^{\alpha_n+2s}, \qquad (5)$$

Z being a Bessel function. The first exponents α_n in each partial series are determined by the conditions (2) to (4) for the functions f_n. In case $kr \ll 1$ we have to use only the lowest term of each partial series, the term with $s = 0$. This means that we may use Laplace's equation $\Delta u = 0$ instead of the wave equation (1).

In the remaining power series the exponents α_n form a sequence of discrete, generally nonrational values from $-\infty$ to $+\infty$. The necessity, however, that u^2 and $(\text{grad } u)^2$ be integrable in the neighborhood of the corner 0, excludes all α_n values below a certain limit. So in (5) a least exponent exists, which we simply call α, the associated f_n being f. In case $kr \ll 1$ we again use only the term with this least exponent and we get with this approximation:

$$u = r^\alpha f(p,\varphi). \qquad (6)$$

The necessity of integrability leads here to the edge condition:

$$\int_o^R (\text{grad } u)^2 \, 2\pi R \, dR \qquad \text{is finite} \qquad (7)$$

(R distance from the edge; Fig. 1)
and to the "corner condition" :

$$\int_o^r (\text{grad } u)^2 \, \pi r^2 \, dr \qquad \text{is finite.}$$

---

† University of Tübingen, Germany.

[1] A. Sommerfeld, *Dtsch. Math. Ver.* vol. 21, pp. 309–357; 1912.
[2] J. Meixner, *Ann. der Phys.*, (6), vol. 6, pp. 2–9; 1949.

The latter gives with (6) :

$$\int_o^r r^{2\alpha}\, dr \qquad \text{is finite}$$

or:

$$\alpha > (-1/2). \tag{8}$$

Thus we have to choose from the discrete sequence of the α_n the least value compatible with the condition (8), unless this value is excluded by the edge condition, which here is a condition for the behavior of $f(p,\varphi)$ in the neighborhood of the edge ($p = p_o$; $\varphi = 0$ or π) ; in the latter case we have to take the next higher value.

So the problem is reduced to the two questions:

1. For each p_o how large is the lowest possible exponent α?

2. What form has the associated function $f(p_o, p, \varphi)$?

METHODS OF SOLUTION

Laplace's equation $\Delta u = 0$ in the co-ordinates r,p,φ assumes the form:

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial u}{\partial r}\right) - 2p\,\frac{\partial u}{\partial p} + (1-p^2)\,\frac{\partial^2 u}{\partial p^2} + (1-p^2)^{-1}\frac{\partial^2 u}{\partial \varphi^2} = 0$$

or with (6):

$$\alpha\,(\alpha+1)\,f - 2p\,\frac{\partial f}{\partial p} + (1-p^2)\,\frac{\partial^2 f}{\partial p^2} + (1-p^2)^{-1}\frac{\partial^2 f}{\partial \varphi^2} = 0. \tag{9}$$

The general solution of this differential equation is:

$$f = n\sum_o^\infty K^{(n)}_\alpha\,(p)\,[C_{1n}\cos n\varphi + C_{2n}\sin n\varphi] \tag{10}$$

The $K^{(n)}_\alpha\,(p)$ are Legendre's associated functions of the first kind of order α, each of which consists of a linear combination of the two linearly independent functions $P^{(n)}_\alpha(p)$ and $Q^{(n)}_\alpha(p)$.[3] As all $P(p)$ diverge for $p = -1$ and all $Q(p)$ for $p = +1$, the solution f has to be composed of two parts and for $p < p_o$ we have to use only the $Q(p)$ for $p > p_o$ only the $P(p)$. The two parts must be determined in such a manner, that f and $\partial f/\partial p$ are continuous at $p = p_o$.

For $p < p_o$ (aperture) $f(\varphi)|_{\rho=\text{const}}$ must be continuous in the whole range $0 < \varphi < 2\pi$, while for $p > p_o$ (screen) it has discontinuities (jumps of $\partial f/\partial \varphi$) at $\varphi = 0$ and $\varphi = \pi$. Thus the series (10), in consequence of the symmetries (2) and (3), can contain only cosine terms with even n for $p < p_o$ and only sine terms with odd n for $p > p_o$. Hence the solution has the form:

$$p < p_o\ ;\ \ f = m\sum_o^\infty C_{2m}\,Q_\alpha^{(2m)}(p)\cos 2m\varphi \tag{11}$$

$$p > p_o\ ;\ \ f = m\sum_o^\infty C_{2m+1}\,P_\alpha^{(2m+1)}(p)\sin\left[(2m+1)\varphi\right], \tag{12}$$

where we define f by (11) and (12) only in the range $0 < \varphi < \pi$, while in the range $-\pi < \varphi < 0$ f is defined by the symmetry (2).

---

[3] See, for instance, W. Magnus and F. Oberhettinger, "Formeln und Sätze für die speziellen Funktionen der mathematischen Physik," 2. Aufl., Berlin, Ger.; 1948.

The differential equation (9) is homogeneous and therefore one constant C, for instance C_o, is arbitrary. The other constants and the possible values of the exponent α are determined by the requirement of continuity of f and $\partial f/\partial p$ at $p = p_o$ for all values of φ.

Suppose for the moment that we know any true exponent α. Then we can compute the $P\,(p_o)$, $Q\,(p_o)$, $\partial P/\partial p]\,(p_o)$ and $\partial Q/\partial p]\,(p_o)$ and the requirement of continuity together with (11) and (12) leads to two equations, which imply the equality of a Fourier series with only cosine terms and a Fourier series with only sine terms in the range $0 < \varphi < \pi$. From these equations we determine approximately the first k coefficients C_1 to C_k (with $C_o = 1$) as follows: we substitute $k/2$ different, for instance equidistant, values of φ (between 0 and $\pi/2$) and get $k/2$ linear equations from each of the two continuity requirements. Thus we obtain k linear equations for the k unknown values C_1 to C_k. This procedure converges with increasing k, if we use a true exponent α in computing the P and Q.

But really we do not know the possible, discrete values of α. They are determined implicitly by even this convergence condition for the procedure just described and cannot be deduced explicitly. So (11) and (12) are indeed the rigorous solution of the problem, but a solution in a useless form: the constants α and C_k cannot be practically determined.

Now we can use quite another method to find a solution of the problem, this time not a rigorous but a practically useful solution. This method depends on the fact that the screen of Fig. 1 becomes a half-plane for $\vartheta_o = \pi/2$, $p_o = 0$, and that for the diffraction at the half-plane we have a rigorous solution by Sommerfeld.[4] We may expand this solution into a series of the form (5), take only the lowest term and have at once $f(p,\varphi)|_{\rho_o=0}$. We find a simple function in a closed form. Now we try to modify this special $f(p,\varphi)$ in such a way that it still satisfies the boundary conditions exactly, now for $p_o \neq 0$, and that it violates the differential equation (9) as little as possible. Later on we determine a "best" value of α for this purpose.

We have much freedom in the way we modify the special $f(p,\varphi)$. We shall try to find a simple, closed expression $f(p_o,p,\varphi)$ which leads us at once to an approximate solution $u(p_o,r,p,\varphi)$ of our problem.

In this procedure we have only to attend to the question, whether the next lower value from the discrete sequence of the α_n—excluded by the requirement of integrability for $p_o = 0$—may not enter the allowed region for $p_o \neq 0$.

THE SPECIAL SOLUTION FOR $p_o = 0$

In case $p_o = 0$ (diffraction at the half-plane) we introduce besides the spherical co-ordinates r,p,φ the cylindrical co-ordinates R,ψ,y (Fig. 1, where for $\vartheta_o = \pi/2$ the edge coincides with the y axis), the connection of

---

[4] A. Sommerfeld, *Math. Ann.*, vol. 47, pp. 317–374; 1896.

which with the spherical co-ordinates is:

$$R = r \left[p^2 + (1 - p^2) \sin^2 \varphi \right]^{1/2}$$
$$\cos \psi = p \left[p^2 + (1 - p^2) \sin^2 \varphi \right]^{-1/2}$$
$$y = r (1 - p^2)^{1/2} \cos \varphi \qquad (13)$$

with all roots positive.

The diffraction wave u (total minus incident wave) for the case of the half-plane, expanded into a series of powers of R, has the form, independent of y [5]:

$$u = C \left\{ (kR)^{1/2} \left| \sin \psi/ 2 \right| + \ldots \ldots (kR)^{3/2} + \ldots . \right\}. \qquad (14)$$

Actually this series in analogy to (5) has the form:

$$u = n \sum g_n(\psi) Z_{\beta_n} (kR) = n \sum g_n(\psi) \sum_{s=0}^{s=\infty} C_s R^{\beta_n + 2s}, \quad (15)$$

with $\beta_n = \ldots . -3/2, -1/2, +1/2, +3/2 \ldots .$.

The terms with negative exponents, however, are excluded by the edge condition (7), which implies:

$$\int_0^R R^{2\beta - 1} \, dR \qquad \text{is finite}$$

or: $\qquad\qquad \beta > 0. \qquad (16)$

The lowest term of (14), transformed into the co-ordinates r, p, φ, becomes:

$$u = 2^{-1/2} C (kr)^{1/2} \left[(\sin^2 \varphi + p^2 \cos^2 \varphi)^{1/2} - p \right]^{1/2} .$$

The comparison with (6) leads for $p = p_o$ to

$$\alpha = 1/2 \quad \text{and (without the constant):}$$
$$f(p,\varphi) = \left[(\sin^2 \varphi + p^2 \cos^2 \varphi)^{1/2} - p \right]^{1/2} . \qquad (17)$$

The exponent $\alpha = 1/2$ is the least one of the sequence $\alpha_n = \ldots . -3/2, -1/2, +1/2, +3/2 \ldots .$ allowed by the corner condition (8). The sequence of the "corner exponents" α_n coincides here (where we have no corner at all) with the sequence of the "edge exponents" β_n, which coincidence does not hold any longer for $p \neq p_o$.

The function (17) is plotted over the p, φ plane in Fig. 2. Exactly the same surface, shifted in the φ direction, is to be thought annexed for the range $-\pi < \varphi < 0$ (or $\pi < \varphi < 2\pi$). This illustrates the continuity of f and $\partial f/\partial \varrho$ for $p < 0$ and the discontinuity along AB and $A'B'$ for $p > 0$. The curves CA and $C'A'$ are parabolas (vertices at A and A'), also the curve DEF (vertex at F). The curve $A\,E\,A'$ is the $(\sin \varphi)^{1/2}$ function.

The function f according to (17) equals zero for $p = +1$ and equals $\sqrt{2}$ for $p = -1$. It satisfies the symmetry conditions (2) and (3), also the boundary condition (4) on the screen. By substituting (17) into (9) one may directly verify the differential equation to be satisfied with $\alpha = 1/2$.

THE SOLUTION FOR $p_o \neq 0$

For $p_o \neq 0$ too, we may represent u by the series (5) in r, p, φ or—but now only as an approximation in the immediate vicinity of the edge, for $R \ll r$—by a series analogous to (15) in r, R, ψ. R and ψ have here the meaning of Fig. 1 and do not satisfy (13) any longer, which are only valid for $p_o = 0$.

[5] W. Braunbek, *Ann. der Phys.*, (6), vol. 6, pp. 53–58; 1949.

We introduce $\varrho = R/r$ and find for $\varrho \ll 1$: Each term of the series (5) (for $s = 0$) may be approximately written:

$$r^{\alpha_n} f_n (p,\varphi) = r^{\alpha_n} \varrho^{\beta_n} g_n(\psi) = r^{\alpha_n - \beta_n} R^{\beta_n} g_n (\psi), \quad (18)$$

and in this way goes over into a term of a series analogous to (15).

Thus to the sequence of corner exponents α_n we may associate a sequence of edge exponents β_n. In case $p_o = 0$ both sequences are identical. Their common least allowed value—selected by the edge condition which is here stronger—is $\alpha = \beta = +1/2$. Turning to the case $p_o \neq 0$, the α_n and the β_n may change continously, but in a different manner.
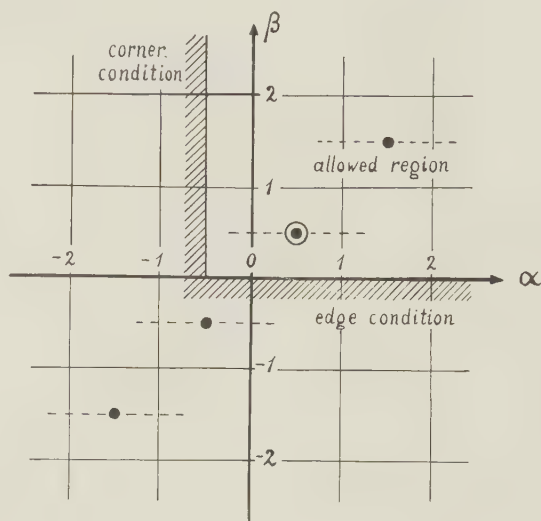
Fig. 2—The function $f(p, \varphi)$ according to (17), plotted over the p, φ plane.

One could think that herewith the next lower pair of values α, β enters the region allowed by the edge condition and the corner condition, or that the lowest pair of values allowed for $p_o = 0$ leaves this region. To answer this, we examine now the change of the exponents β_n with changing p_o.

The new co-ordinates ϱ, ψ have (always for $\varrho \ll 1$!) the following connection with p, φ:

$$\varrho \cos \psi = (1 - p_o^2)^{-1/2} (p - p_o)$$
$$\varrho \sin \psi = (1 - p_o^2)^{1/2} \varphi . \qquad (19)$$

If by this relations we transform the differential equation (9) into the variables ϱ, ψ, take (18) and use only the terms with the lowest power of ϱ, we get:

$$\beta^2 g + \frac{d^2 g}{d\psi^2} = 0$$

with the partial solutions:

$$g(\psi) = C \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} (\beta\psi). \qquad (20)$$

The boundary condition (4) allows only sine functions, the symmetry condition (2) only the values $\beta = \cdots -3/2, -1/2, +1/2, +3/2 \cdots$. Thus we find that the sequence of the β for all p_o remains the same as for $p_o = 0$. This result occurs, because all terms in the differential equation containing p_o or $\alpha(p_o)$ have higher powers of ϱ and so are dropped for $\varrho \ll 1$.

Plotting the possible pairs of values α, β in a diagram (Fig. 3), which shows also the allowed region and its limits according to the edge condition and the corner condition, we perceive that—with the (α, β) points moving only horizontally—no lower point can enter the allowed region. Furthermore the lowest point allowed at $p_o = 0$, the point $(1/2, 1/2)$, does not leave this region unless its α value diminishes at least to $(-1/2)$. This case does not occur as will be shown later on.
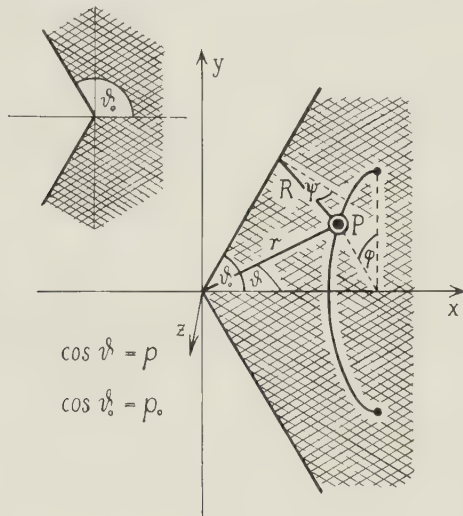
Fig. 3—The possible pairs of values α, β and the region allowed by the edge condition and the corner condition.

Thus at arbitrary p_o that α value remains the lowest allowed, which continously changes from $\alpha = +1/2$ at $p_o = 0$. Exactly the same result would follow without taking into consideration the vertical limit in Fig. 3 at all. This means that the corner condition has no actual influence and that therefore it is not necessary to introduce a corner condition besides the well-known edge condition.

Now we may continue in getting approximately the exponent $\alpha(p_o)$ and the function $f(p_o, p, \varphi)$ for an arbitrary p_o. From (20) comes:

$$g(\psi) \sim \sin \psi/2 \sim (1 - \cos \psi)^{1/2}$$

or with (18) and (19):

$$f(p_o, p, \varphi) \sim \varrho^{1/2}(1 - \cos \psi)^{1/2}$$

$$\sim \left[\left(\varphi^2 + \left(\frac{p - p_o}{1 - p_o^2} \right)^2 \right)^{1/2} - \frac{p - p_o}{1 - p_o^2} \right]^{1/2} . \quad (21)$$

This holds only for $\varrho \ll 1$; i.e., $|p - p_o| \ll 1$ and $\varphi \ll 1$. In case $p_o = 0$ it assumes the form:

$$f(0, p, \varphi) \approx [(\varphi^2 + p^2)^{1/2} - p]^{1/2}$$

identical with (17) under the restrictions above mentioned. A comparison of (21) and (17) shows how to generalize the expression (17) to the case $p_o \neq 0$, without losing the true singularity in the neighborhood of the edge. The simplest possible form of such a generalization is:

$$f(p_o, p, \varphi) = [(\sin^2 \varphi + F(p_o, p) \cos^2 \varphi)^{1/2} - F(p_o, p)]^{1/2}, \quad (22)$$

with $F(p_o, p)$ satisfying the following conditions:

$$\left. \begin{array}{l} F(p_o, -1) = -1 \\ F(p_o, +1) = +1 \\ \lim_{\rho o \to o} F = p \\ \lim_{\rho \to \rho o} F = \dfrac{p - p_o}{1 - p_o^2} \\ \dfrac{\partial F}{\partial p} < 0 \end{array} \right\} \quad (23)$$

for all p_o and p between -1 and $+1$.

The last of these conditions is established only intuitively by considering the deformation of the surface Fig. 2 and of the curve DEF when we shift the singular points A and A' in the direction to B and B' or in the opposite direction.

The most simple function satisfying the conditions (23) is:

$$F(p_o, p) = \frac{p - p_o}{1 - p_o p} . \quad (24)$$

This function does not contain any free constant which could be adjusted to the problem. Therefore we can only expect a rather poor approximation with this function. We could find more complicated functions instead of (24) or even already instead of (22), which also satisfy the conditions (23) and which would allow for adjusting free constants. By this we could hope to satisfy the differential equation (9) better and better and so to find a better and better approximation.

In the present paper, however, we shall proceed only with the expressions (22) and (24). Substituting these functions into the differential equation (9), which contains the exponent α yet unknown, we can try to find the "best" value of α, which makes the deviation of the differential expression $\Phi(\alpha, p_o, p, \varphi)$ on the left hand of (9) from zero a minimum. We could require:

$$\int_{p=-1}^{p=+1} \int_{\varphi=0}^{\varphi=\pi} \Phi^2 \, d\varphi \, dp = \text{Min}$$

or:

$$\frac{\partial}{\partial \alpha} \iint \Phi^2 \, d\varphi \, dp = 0 ,$$

from which we find $\alpha(p_o)$.

As the expression Φ is rather complicated and the computing of the double integral is rather troublesome, we rest content with the more simple claim of satisfying the differential equation only along the curve DEF ($\varphi = \pi/2$; Fig. 2) or along the curve CA ($\varphi = 0$) as well as possible. This leads to:

$$\frac{\partial}{\partial \alpha} \int \Phi^2_{\varphi = \text{const.}} \, dp = 0 \quad (25)$$

with $\pi/2$ or 0 for the constant value of φ.

Substituting (22) and (24) into the differential equation (9), we find as a result of a longer computation:

$$\Phi_{\varphi=\pi/2} = (1 + p_o)^{1/2}(1 - p)^{1/2}(1 - p_o p)^{-5/2}[\alpha(\alpha + 1) (1 - p_o p)^2 - \tfrac{3}{4}(1 - p_o^2)]$$

$$\Phi_{\varphi=0} = 2^{1/2}(p_o - p)^{1/2}(1 - p_o p)^{-5/2}[\alpha(\alpha + 1) (1 - p_o p)^2 - \tfrac{3}{4}(1 - p_o^2)] \quad \text{for } p < p_o .$$

In case $\varphi = \pi/2$ (integration along DEF from $p = -1$ to $p = +1$), (25) gives:

$$\gamma = \alpha(\alpha + 1) = 3p_o^2$$
$$\left[4p_o(1 + p_o) + 2(1 - p_o^2) \, ln \, \frac{1 - p_o}{1 + p_o} \right]^{-1} , \qquad (26)$$

and for $\varphi = 0$ (integration along CA from $p = -1$ to $p = p_o$):

$$\gamma' = \alpha'(\alpha' + 1) = 3p_o^2[8p_o(1 + p_o) + 8(1 - p_o^2)$$
$$ln(1 - p_o)]^{-1} . \qquad (27)$$

From both γ or γ' we find the value of α or α':

$$\alpha = (\gamma + 1/4)^{1/2} - 1/2 . \qquad (28)$$

As the two formulas (26) and (27) represent two different and rather poor approximations to the true function $\gamma(p_o)$, we may not expect that they lead to the same result. The difference between γ and γ', or between α and α', can, on the other hand, supply a measure of the inaccuracy of this rather rough simplification of the general method.

With $p_o \rightarrow 0$ the square brackets in (26) and (27) approach zero as $4\,p_o^2$ and we get in accordance with the former result:

$$\gamma(0) = \gamma'(0) = 3/4$$
$$\alpha(0) = \alpha'(0) = 1/2 .$$

The functions $\alpha(p_o)$ and $\alpha'(p_o)$ according to (26), (27) and (28) are plotted in Fig. 4. The difference of the two functions is considerable. The true function $\alpha(p_o)$ may lie between the two curves of Fig. 4. At all events we see that

$$\alpha < 1/2 \text{ for } p_o > 0 \text{ (salient corner)}$$
$$\alpha > 1/2 \text{ for } p_o < 0 \text{ (re-entering corner)}.$$

The total function $u(p_o, r, p, \varphi)$ is at this stage of approximation, according to (6), (22) and (24):

$$u = \gamma^\alpha \left[(\sin^2 \varphi + \left(\frac{p - p_o}{1 - p_o p} \right)^2 \cdot \cos^2 \varphi)^{1/2} - \frac{p - p_o}{1 - p_o p} \right]^{1/2} . \qquad (29)$$

We may here add the interesting consequence, that in the case of a re-entering corner the exponent α (Fig. 4) necessarily traverses unity (at $p_o \sim -0.7$; $\vartheta_o \sim 135$ degrees). At this point the gradient $(\partial u/\partial r)_o$ at the corner point jumps with decreasing p_o from infinity to zero.
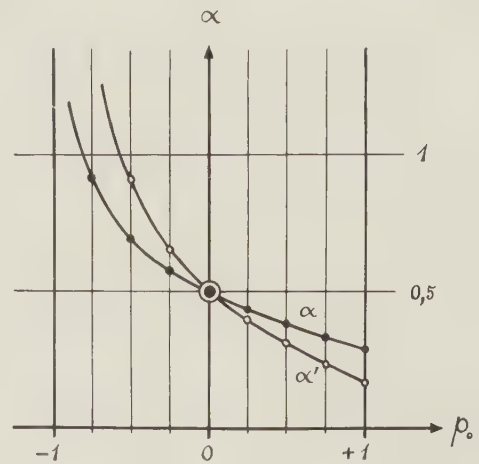
Fig. 4—The dependence of the exponent α on the parameter p_o. α according to (26), α' according to (27).

Naturally it is not possible by the method used here to determine the multiplicative constant C belonging to $f(p_o, p, \varphi)$, as this constant is connected with the amplitude of the incident wave and cannot be determined from an approximation confined to the range $kr \ll 1$.

After this work was done, it was called to the author's attention that a similar problem with electromagnetic waves was treated by Jones.[6] Jones finds in part similar results, especially the superfluity of a separate corner condition.

[6] D. S. Jones, *Quart. Jour. Mech. and Appl. Math.*, vol. V, pt. 3' 1952.

Electromagnetic Radiation Patterns and Sources*

CLAUS MÜLLER

I. Introduction

CONSIDER an electromagnetic field described by the vectors E and H which satisfy the equations (ω being a positive and real constant)

$$\nabla \times H + i\omega E = 0 \qquad \nabla \times E - i\omega H = 0 \qquad (1)$$

outside a finite region G. If this field satisfies the radiation conditions

$$n \times H = -E + o\left(\frac{1}{r}\right); E = 0\left(\frac{1}{r}\right), \qquad (2)$$

it may be regarded as being generated by sources all of which are inside G.

The radiation pattern F is then given by the asymptotic behavior of the field E, H at infinity. In order to formulate the properties of the patterns we introduce spherical co-ordinates

$$(x) = r(\xi) \qquad (3)$$

where (ξ) is the unit direction vector of the co-ordinate vector (x).

Our first result is the proof that any field E, H which has the properties mentioned above satisfies an asymptotic expansion

$$E = \frac{e^{i\omega r}}{r} F(\xi) + o\left(\frac{1}{r}\right); H = \frac{e^{i\omega r}}{r} n \times F(\xi) + o\left(\frac{1}{r}\right), \quad (4)$$

where the vector field $F(\xi)$, which will be called the radiation pattern of the electromagnetic field E, H, can be determined and satisfies the condition

$$(nF(\xi)) = 0. \qquad (5)$$

The radiation pattern is therefore a complex valued vector field, defined on the unit sphere, which has no components in the direction of the normal.

By establishing necessary and sufficient conditions for a field $F(\xi)$ to be a pattern we obtain a full description of the asymptotic behavior of the electromagnetic radiation fields. In other words, we can determine whether for a given vector field $F(\xi)$ there exists an electromagnetic field E, H such that the asymptotic relation (4) is satisfied.

These results are extensions of an earlier paper[1] on the corresponding problem for the scalar Helmholtz equation. We had proved there, that if $U(x)$ is a solution of

---

[1] C. Muller, "Radiation Patterns and Radiation Fields," New York University, Inst. Mat. Sci., Div. Electromagnetic Res., Res. Rep. No. EM–62, March, 1954, or *Jour. Rat. Mech. and Anal.*, vol. 4, p. 235; 1955.

$$\Delta U + \omega^2 U = 0 \qquad (6)$$

for $|x| \geq R$, which satisfies the Sommerfeld radiation condition,

$$\frac{\partial U}{\partial n} = i\omega U + o\left(\frac{1}{r}\right), \qquad (7)$$

uniformly for all directions, then there is a function $f(\xi)$ such that

$$U(x) = \frac{e^{i\omega r}}{r} f(\xi) + 0\left(\frac{1}{r^2}\right). \qquad (8)$$

The function $f(\xi)$ was called the (scalar) radiation pattern of $U(x)$. Our previous result may be summed up in

Theorem 1: A necessary and sufficient condition that a function $f(\xi)$ defined on the unit sphere be a radiation pattern, is that there exist an entire harmonic function $h(x)$ with the following properties:

1. $\Delta h(x) = 0$ for all (x)

2. $h(x) = f(\xi)$ for $(x) = (\xi)$

3. $\overline{\lim_{r \to \infty}} \frac{1}{\omega r} \log \int_{|x|=r} |h(r\xi)|^2 \, dS = R < \infty.$

If we formulate (1) and (2) in terms of E above we get

$$\nabla \times \nabla \times E = \omega^2 E \qquad (9a)$$

$$n \times (\nabla \times E) = -i\omega E + o\left(\frac{1}{r}\right) \qquad (9b)$$

$$E = o\left(\frac{1}{r}\right). \qquad (9c)$$

These relations will be shown to be equivalent to

$$\Delta E + \omega^2 E = 0 \qquad (10a)$$

$$\frac{\partial}{\partial n} E = i\omega E + o\left(\frac{1}{r}\right) \qquad (10b)$$

$$(nE) = o\left(\frac{1}{r}\right). \qquad (10c)$$

As (9b, 9c) and (10b, 10c) were both used as radiation conditions this result settles the question of the validity of the two conditions.

In order to show that (10) follows from (9) we express E by an integral over a large but finite sphere and then prove (10) explicitly. To deduce (9) from (10) is more interesting, as the essential argument shows that $\nabla E = 0$ follows from the asymptotic property

$$(nE) = o\left(\frac{1}{r}\right).$$

We may regard this as another instance of the importance of the asymptotic behavior for the solutions of the Helmholtz equation.

We give the exact formulation in

Lemma 1: Suppose E is a vector field which satisfies

$$\Delta E + \omega^2 E = 0 \text{ and } \frac{\partial}{\partial n} E = i\omega E + o\left(\frac{1}{r}\right)$$

for $|x| \geq R$. Then if

$$(En) = o\left(\frac{1}{r}\right)$$

for $r \rightarrow \infty$, it follows that for $|x| < R$

$$\nabla E = 0.$$

To prove this we form the scalar function

$$V(x) = (x \cdot E) \tag{11}$$

where $(x \cdot E)$ denotes the scalar product of the co-ordinate vector (x) and the vector field E. Since

$$\Delta V = (x \cdot \Delta E) + 2\nabla \cdot E \tag{12}$$

we get

$$(\Delta + \omega^2)V = 2\nabla \cdot E \tag{13}$$

and

$$(\Delta + \omega^2)^2 \, V = 0. \tag{14}$$

The function V therefore is a solution of the iterated Helmholtz equation, and

$$V(x) = o(1)$$

for $r \rightarrow \infty$. From this behavior we shall show later that

$$(\Delta + \omega^2)V = 0 \tag{16}$$

and thus prove Lemma 1.

The equivalence of (9) and (10) makes it possible now to determine the properties of the vector radiation patterns by using Theorem 1 for each of the three Cartesian co-ordinates of E. We then get

Theorem 2: A necessary and sufficient condition that a vector field $F(\xi)$, defined on the unit sphere be a vector radiation pattern, is that there exist an entire harmonic vector field $H(x)$ with the following properties:

1. $\Delta H(x) = 0$ for all (x)
2. $H(x) = F(\xi)$ for $(x) = (\xi)$
3. $(xH(x)) = 0$ for $|x| = 1$
4. $\overline{\lim_{r \to \infty}} \dfrac{1}{\omega r} \log \displaystyle\int_{|x|=r} |H(r\xi)|^2 \, dS = R < \infty.$

After having determined the properties of the vector radiation patterns we shall proceed to discuss the relation of the sources and the patterns. We shall confine ourselves to the case of electric sources only, which may be isolated electric dipoles or continuous distributions of dipole densities on lines, surfaces and in domains. In the latter cases we may speak of "currents." We shall first see that if we do not impose further restrictions the sources are not uniquely determined by the pattern.

To see this let us assume that to a given pattern $F(\xi)$ we constructed the vector field E, H for $|x| > R$. The field E is analytic in this region and is therefore twice continuously differentiable on the sphere $|x| = R_1$, $R_1 > R$. We may then continue the field E to the inside of $|x| = R_1$ in such a way that E is twice continuously differentiable everywhere. There are many possibilities of doing this and the process of continuation is not at all uniquely determined. For $|x| \leq R_1$ we may form

$$\nabla \times \nabla \times E - \omega^2 E = -\frac{i}{\omega} 4\pi j. \tag{17}$$

This vector field j then gives us electric volume currents which generate the field E, H. In order to prove this we set

$$H = -\frac{i}{\omega} \nabla \times E \tag{18}$$

which then satisfies the equation

$$\nabla \times E - i\omega H = 0 \tag{19}$$

everywhere. From (17) we get

$$\nabla \times H + i\omega E = \begin{cases} 0 \text{ for } |x| > R_1 \, . \\ j \text{ for } |x| \leq R_1 \end{cases} \tag{20}$$

Using the vector Green's formula for our equations we then get a representation of E and H in terms of the electric volume currents j.

However these currents j are not uniquely determined and we may find many different current distributions according to the different continuations. The volume currents are therefore not uniquely determined and we must impose further conditions on our currents if we want to establish a one-to-one relation between the patterns and the volume currents. In order to find these conditions let us assume first, that there are two different currents distributions j_1 and j_2 in a region G which generate the same pattern. The corresponding fields are therefore identical outside G, and the difference $j_1 - j_2 = j_0$ represents a current density generating a field which vanishes everywhere outside G. To any current j we may now add the field j_0 without changing the radiation pattern.

If we want to find the most "efficient" current corresponding to a given pattern it seems natural to require that the integral

$$\int_G |j|^2 \, dV \tag{21}$$

be as small as possible. By introducing this auxiliary condition we will later be able to determine the currents uniquely. However, before we can do this we have to study the idea of the efficiency of a distribution of sources more closely (Section III).

We can best do this by discussing the simple case of two dipoles j_1 and j_2 which are fixed at two points (x_1) and (x_2). Let e_1, e_2 and e_3 be an orthonormal basis of a system of Cartesian co-ordinates. We may then put

$$j_1 = j_1{}^i e_i \; ; j_2 = j_2{}^i e_i \, , \tag{22}$$

assuming the usual convention of summation. The dipoles e_i when treated at (x_1) will generate radiation patterns $P_i{}^1(\xi)$. If these dipoles are at (x_2) they lead correspondingly to patterns $P_i{}^2(\xi)$. Any radiation pattern which we get from two dipoles at (x_1) and (x_2) is therefore of the form

$$j_1{}^i P_i{}^1(\xi) + j_2{}^i P_i{}^2(\xi) = F(\xi). \tag{23}$$

The total radiation of this pattern is

$$\lim_{r \to \infty} \int_{|x|=r} n(\overline{E} \times H) \, dS = C. \tag{24}$$

From (2) we get

$$C = \lim_{r \to \infty} \int_{|x|=r} E\overline{E} \, dS, \qquad (25)$$

and (4) leads to

$$C = \int_{|\xi|=1} |F(\xi)|^2 \, dS. \qquad (26)$$

If $j_2 = 0$ we can easily evaluate this integral and obtain

$$C_1 = \frac{8\pi}{3} \omega^2 j_1 \bar{j}_1. \qquad (27)$$

Correspondingly, for $j_1 = 0$ we would get

$$C_2 = \frac{8\pi}{3} \omega^2 j_2 \bar{j}_2. \qquad (28)$$

Now the total radiation of the two dipoles together is in general not equal to the sum of their individual radiations, as the interaction of the two fields will either decrease or increase the total radiation. This implies, of course, that we have assumed that the two dipoles oscillate coherently. If we had assumed that they oscillate incoherently the total radiation would be $C_1 + C_2$ since the interaction vanishes if we average over a random distribution of phases.

We now define

$$W = \text{Efficiency} = \frac{\text{Total coherent radiation}}{\text{Total incoherent radiation}} = \frac{C}{I}.$$

From (23) and (26) we get

$$C = \int_{|\xi|=1} |F(\xi)|^2 \, dS = \int_{|\xi|=1} |j_1{}^i P_{\cdot}{}^1(\xi) + j_2{}^i P_{\cdot}{}^2(\xi)|^2 \, dS. \qquad (29)$$

This is a Hermitian form in the six variables $j_1{}^i$ and $j_2{}^i$. Introducing the vector

$$J = (j_1{}^1, j_1{}^2, j_1{}^3, j_2{}^1, j_2{}^2, j_2{}^3), \qquad (30)$$

we may write (29) in the form

$$C = \frac{8\pi}{3} \omega^2 (J, M\overline{J}), \qquad (31)$$

where M is a Hermitian matrix of six rows and columns. A detailed analysis shows that M is real and symmetric. In this notation,

$$I = \frac{8\pi}{3} \omega^2 |J|^2 \qquad (32)$$

and we find

$$W = \frac{(J, M\overline{J})}{J\overline{J}}. \qquad (33)$$

The full discussion of the patterns generated by two dipoles may thus be carried out by determining the six eigenvectors and eigenvalues of the matrix C. Each of these eigenvectors J_k then represents a certain arrangement of dipoles, the most efficient of which is given by the eigenvector with the greatest eigenvalue.

The patterns corresponding to these eigenvectors form an orthogonal basis for all the patterns which may be generated by two dipoles in fixed locations.

This method is not limited to the case of two dipoles but may be extended to describe the patterns of any finite number of dipoles in fixed points. If the number of dipoles is N, this leads to a symmetric and real matrix of $3N$ rows and determines an orthogonal basis of $3N$ patterns.

The case of the two dipoles is discussed in detail, giving the eigenvalues and eigenvectors (Section III).

We can proceed to continuous dipole distributions on lines, surfaces and in regions. Here again we have to define the efficiency and determine a basis of the possible patterns.

The procedure is a formal extension of the arguments which we used for the discrete dipoles.

For an arbitrary current distribution j we calculate the total radiation, which can be written as

$$\frac{8\pi}{3} \omega^2 \iint \left(j(x), M(x, y)\overline{j(y)} \right) d\tau_x \, d\tau_y = C, \qquad (34)$$

where the matrix $M(x, y)$ may be expressed as

$$M(x, y) = \left[\left(\psi_o(\omega|x - y|) - \frac{1}{2} \psi_2 (\omega|x - y|) \right) \delta_{ik} + \frac{3}{2} \frac{\psi_2(\omega|x - y|)}{|x - y|^2} (x - y)_i (x - y)_k \right] \qquad (35)$$

and the double integration is taken over the domain in which we are interested, $d\tau_x$ and $d\tau_y$ representing the respective elements of integration.

We now define I by

$$\frac{8\pi}{3} \omega^2 \int |j(x)|^2 \, d\tau_x = I, \qquad (36)$$

and put

$$W = \frac{C}{I} = \frac{\iint \left[j(x), M(x, y)\overline{j(y)} \right] d\tau_x \, d\tau_y}{\int |j(x)|^2 \, d\tau_x}. \qquad (37)$$

The matrix M is the kernel of a symmetric system of three integral equations which we write compactly in vector form as

$$j_k(x) = \lambda_k \int M(x, y) j_k(y) \, d\tau_y. \qquad (38)$$

These vector fields j_k are real and may be normalized in such a way that

$$\frac{8\pi}{3} \omega^2 \int_G j_k \cdot j_l \, dV = \delta_{kl}. \qquad (39)$$

The most interesting use of these ideas is perhaps made by establishing a relation between the radiation patterns and the distribution of sources in a given region G.

To this end we start with the vector fields $j_k(x)$ which solve (38). Each of these vector fields generates a pattern $F_k(\xi)$. If $F_l(\xi)$ and $F_k(\xi)$ are two of these patterns, the current $aj_k + bj_l$ will yield the pattern $aF_k(\xi) + bF_l(\xi)$. The total radiation of this current distribution thus is

$$\int_{|\xi|=1} |aF_k(\xi) + bF_l(\xi)|^2 \, dS_\xi$$

$$= \frac{8\pi}{3} \omega^2 \int_G \int \left(\left[aj_k(x) + bj_l(x) \right], \right.$$

$$M(x, y) \left[\overline{aj_k(y)} + \overline{bj_l(y)} \right] \right) dV_x \, dV_y$$
$$= (\lambda_k |a|^2 + \lambda_l |b|^2) \qquad (42)$$

where the last equation follows from (38). By comparing coefficients we get, therefore,

$$\int_{|\xi|=1} F_k(\xi) \cdot \overline{F_l(\xi)} \, dS_\xi = \frac{1}{\lambda_k} \delta_{kl} = W_k \delta_{kl} . \qquad (43)$$

Suppose now that $F(\xi)$ is a radiation pattern, which can be generated by sources all of which are inside of G. Then the corresponding field E, H is analytic on the boundary S of G. We may therefore continue the field E to the inside of G in such a way that it is twice continuously differentiable everywhere. By forming

$$\nabla \times \nabla \times E - \omega^2 E = - \frac{i}{\omega} 4\pi j \qquad (44)$$

inside of G we thus get a current j which generates $F(\xi)$. Now take the expansion coefficients with respect to the patterns $F_k(\xi)$.

The vector fields j_k are real and may be normalized so that

$$\frac{8\pi}{3} \omega^2 \int_G j_k(x) j_l(y) \, dV_x = \delta_{kl} . \qquad (45)$$

Then the coefficients

$$\frac{8\pi}{3} \omega^2 a_k = \int_G j(x) \cdot j_k(x) \, dV_k \qquad (46)$$

satisfy the relation

$$\frac{8\pi}{3} \omega^2 \sum_{k=1}^{\infty} |a_k|^2 \leq \int_G |j(x)|^2 \, dV_x . \qquad (47)$$

Since the $j_k(x)$ do not represent a complete set, the equality sign will not hold in general. The total radiation of the sources j now is equal to

$$\frac{8\pi}{3} \omega^2 \iint_G \left(j(x), M(x, y) \overline{j(y)} \right) dV_x \, dV_y , \qquad (48)$$

and thus becomes

$$\sum_{k=1}^{\infty} |a_k|^2 W_k . \qquad (49)$$

If $F(\xi)$ is the desired pattern, we form the currents

$$j_N = \sum_{k=1}^{N} a_k j_k(x). \qquad (50)$$

They generate the pattern

$$F_N(\xi) = \sum_{k=1}^{N} a_k F_k(\xi). \qquad (51)$$

The total radiation of the current $j - j_N$ therefore is

$$\iint_G \left[(j(x) - j_N(x)), M(x, y) \left(\overline{j(y)} - \overline{j_N(y)} \right) \right] dV_x \, dV_y$$
$$= \frac{3}{8\pi \omega^2} \sum_{k=N+1}^{\infty} |a_k|^2 W_k , \qquad (52)$$

which is equal to

$$\int_{|\xi|=1} |F(\xi) - F_N(\xi)|^2 \, dS_\xi . \qquad (53)$$

Since the right hand side of (52) tends towards zero

as N goes to infinity, we have

$$\lim_{N \to \infty} \int_{|\xi|=1} |F(\xi) - F_N(\xi)|^2 \, dS_\xi = 0. \qquad (54)$$

We now introduce the normalized patterns

$$\sqrt{W_k} \, P_k(\xi) = F_k(\xi) \qquad (55)$$

as a basis for all those patterns which can be generated by sources all of which are inside G. Then

$$F_N(\xi) = \sum_{k=1}^{N} \sqrt{W_k} a_k P_k(\xi) \qquad (56)$$

and

$$\int_{|\xi|=1} |F(\xi) - F_N(\xi)|^2 \, dS_\xi = \int_{|\xi|=1} |F(\xi)|^2 \, dS_\xi$$
$$- \sum_{k=1}^{N} \sqrt{W_k} (a_k \overline{b_k} + \overline{a_k} b_k) + \sum_{k=1}^{N} W_k a_k \overline{a_k} , \qquad (57)$$

with

$$b_k = \int_{|\xi|=1} F(\xi) \cdot \overline{P_k(\xi)} \, dS_\xi . \qquad (58)$$

The right hand side of (57) can only tend towards zero for $N \to \infty$, if

$$b_k = \sqrt{W_k} \, a_k \qquad (59)$$

because these values b_k give the minimum value of (57). From (53) it follows therefore that the $P_k(\xi)$ are an orthonormal basis for our patterns.

From (47) we have now

$$\sum_{k=1}^{\infty} \frac{|b_k|^2}{W_k} < \infty . \qquad (60)$$

This gives us

Theorem 3: A necessary condition for the pattern $F(\xi)$ to be representable by sources all of which are inside the region G is that the expansion coefficients

$$b_k = \int_{|\xi|=1} F(\xi) \overline{P_k}(\xi) \, dS_\xi$$

satisfy

$$\sum_{k=1}^{\infty} \frac{|b_k|^2}{W_k} < \infty .$$

It may be stated that this condition is also sufficient, though we shall not give the details of the proof.

The following sections give first the proof of Theorem 2 and then discuss the concepts introduced in the foregoing discussion for the case of the isolated dipoles.

II. The Electromagnetic Radiation Patterns

We suppose that E, H is a vector field which is a solution of the equations

$$\nabla \times H + i\omega E = 0; \qquad \nabla \times E - i\omega H = 0 \qquad (61)$$

for $|x| \geq R > 0$ and which satisfies the radiation conditions

$$n \times H = -E + o\left(\frac{1}{r}\right); \qquad E = o\left(\frac{1}{r}\right) \qquad (62)$$

for $r \to \infty$, uniformly with respect to all directions. We form the field

$$E' = \begin{cases} E \text{ for } |x| \geq R_1 > R \\ 0 \text{ for } |x| < R_1 \end{cases}$$

$$H' = \begin{cases} H \text{ for } |x| \geq R_1 > \text{R} \\ 0 \text{ for } |x| < R_1 \end{cases}. \tag{63}$$

This field is a solution of Maxwell's equations everywhere except on $|x| = R_1$ where the tangential components have the jumps

$$n \times E_e' - n \times E_i' = n \times E$$
$$n \times H_e' - n \times H_i' = n \times H \tag{64}$$

the subscripts e and i denoting the limits when the field approaches the surface $|x| = R_1$ from the exterior or from the interior.

Let us now regard the field

$$E^* = \frac{1}{4\pi} \int_{|x|=R_1} \left[i\omega j\phi - j' \times \nabla\phi + \rho\nabla\phi \right] dF,$$

$$H^* = \frac{1}{4\pi} \int_{|x|=R_1} \left[i\omega j'\phi + j \times \nabla\phi + \rho'\nabla\phi \right] dF, \tag{65}$$

which is generated by the electric and magnetic surface currents j and j' and the associated charges ρ and ρ'. This field again satisfies Maxwell's equations everywhere except on $|x| = R_1$, where we have the jumps

$$n \times E_e^* - n \times E_i^* = -j',$$
$$n \times H_e^* - n \times H_i^* = j, \tag{66}$$

provided j and j' are continuous and ρ and ρ' are Holder continuous. Now if we set

$$j' = -n \times E; \qquad j = n \times H, \tag{67}$$

then these conditions are satisfied and the fields E', H' and E^*, H^* have the same jumps. Therefore their difference E_o, H_o is a solution of Maxwell's equation for all x with $|x| \neq R_1$. The tangential components of (E_o, H_o) have the same values when approaching $|x| = R_1$ from either side. As E_o, H_o satisfies the radiation conditions, we get, from the uniqueness theorem, that this field vanishes identically.

For $|x| > R_1$ we may thus represent E, H by the surface integrals given in (65). These integrals now show immediately that each of the three Cartesian co-ordinates of E or H satisfies the Sommerfeld radiation conditions.

From (61) we get

$$\nabla \times \nabla \times E = \omega^2 E \tag{68}$$

and (62) is equivalent to

$$n \times \nabla \times E = -i\omega E + o\left(\frac{1}{r}\right). \tag{69}$$

It follows from (68) and (69) therefore that

$$(\Delta + \omega^2)E = 0 \text{ for } |x| \geq R_1 > R > 0 \tag{70}$$

and

$$\frac{\partial}{\partial n} E - i\omega E = o\left(\frac{1}{r}\right) \text{ and } E \cdot n = o\left(\frac{1}{r}\right). \tag{71}$$

These two properties imply according to Theorem 1 which was stated in the Introduction that there is a vector field $F(\xi)$ defined on the unit sphere $|\xi| = 1$ such that

$$E = \frac{e^{i\omega r}}{r} F(\xi) + o\left(\frac{1}{r}\right) \tag{72}$$

and

$$nF(\xi) = 0. \tag{73}$$

This vector field $F(\xi)$ then has the properties formulated in Theorem 2. We have thus proved that this property is a necessary condition for $F(\xi)$ to be a radiation pattern.

In order to show that this condition is sufficient we have now to deduce (68) and (69) from (70) and (71).

To this end we use

Lemma 2: Suppose the scalar function U is a solution of the iterated equation

$$(\Delta + \omega^2)^2 U = 0$$

for $|x| \geq R > 0$. Then if

$$U = o(1)$$

for $|x| \to \infty$ uniformly with respect to all directions, the function U satisfies the noniterated equation

$$(\Delta + \omega^2)U = 0.$$

The proof requires a system $S_{n,j}(\xi)$ of orthonormal spherical harmonics. We form

$$A_{n,j}(r) = \int_{|\xi|=1} U(r\xi)S_{n,j}(\xi) \, dS_\xi. \tag{74}$$

Then for $r > R$

$$\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} A_{n,j}(r) = \int_{|\xi|=1} (\Delta U - \frac{1}{r^2}\Delta_o U)S_{n,j}(\xi) \, dS, \tag{75}$$

where Δ_o is the Beltrami's operator on the unit sphere. Since

$$\Delta_o S_{n,j}(\xi) + n(n+1)S_{n,j} = 0, \tag{76}$$

we have

$$\int_{|\xi|=1} S_{n,j}\Delta_o U(r\xi) \, dS_\xi = \int_{|\xi|=1} U(r\xi)\Delta_o S_{n,j} \, dS_\xi$$

$$= -n(n+1) \int_{|\xi|=1} U(r\xi)S_{n,j} \, dS = -n(n+1)A_{n,j}, \tag{77}$$

and thus obtain

$$\left(\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} + \omega^2 - \frac{n(n+1)}{r^2}\right)A_{n,j}$$

$$= \int_{|\xi|=1} S_{n,j}(\xi)(\Delta + \omega^2)U \, dS. \tag{78}$$

Applying this process again we get

$$\left(\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} + \omega^2 - \frac{n(n+1)}{r^2}\right)^2 A_{n,j}$$

$$= \int_{|\xi|=1} S_{nij}(\xi)(\Delta + \omega^2)^2 U \, dS = 0. \tag{79}$$

The coefficients $A_{n,j}$, therefore, are solutions of this homogeneous fourth-order differential equation.

Set

$$\zeta_n^{(1)} = \sqrt{\frac{2}{\pi r}} H^{(1)}_{(2n+1)/2}{}^{(1)}(r);$$

$$\zeta_n^{(2)}(r) = \sqrt{\frac{2}{\pi r}} H^{(2)}_{(2n+1)/2}(r) \tag{80}$$

and

$$\chi_n^{(1)}(r) = \frac{1}{2i} \int_1^r \rho^2 \zeta_n^{(1)}(\rho) \left[\zeta_n^{(1)}(r)\zeta_n^{(2)}(\rho) \right. $$
$$\left. - \zeta_n^{(2)}(r)\zeta_n^{(1)}(\rho) \right] d\rho$$

$$\chi_n^{(2)}(r) = \frac{1}{2i} \int_1^r \rho^2 \zeta_n^{(2)}(\rho) \left[\zeta_n^{(1)}(r)\zeta_n^{(2)}(\rho) \right. \quad (81)$$
$$\left. - \zeta_n^{(2)}(r)\zeta_n^{(1)}(\rho) \right] d\rho.$$

Then

$$\left[\frac{1}{r^2} \frac{d}{dr} r^2 \frac{d}{dr} + \left(1 - \frac{n(n+1)}{r^2} \right) \right] \chi_n^{(k)}(r)$$
$$= \zeta_n^{(k)}(r) \qquad (k = 1, 2). \quad (82)$$

Therefore these four functions represent a basis for the solutions of the differential equation (79), if $\omega = 1$.

From the asymptotic expansions

$$\zeta_n^{(1)}(r) = e^{i(r-\pi n/2)}/r + O\left(\frac{1}{r^2}\right),$$

$$\zeta_n^{(2)}(r) = e^{-i(r-\pi n/2)}/r + O\left(\frac{1}{r^2}\right), \quad (83)$$

we get after an elementary calculation
$$2i\chi_n^{(1)}(r) = e^{i(r-\pi n/2)} + o(1)$$
$$2i\chi_n^{(2)}(r) = - e^{-i(r-\pi n/2)} + o(1). \quad (84)$$

It now follows from (79) that
$$A_{n,j}(r) = a_{n,j}\zeta_n^{(1)}(\omega r) + b_{n,j}\zeta_n^{(2)}(\omega r)$$
$$+ c_{n,j}\chi_n^{(1)}(\omega r) + d_{n,j}\chi_n^{(2)}(\omega r), \quad (85)$$
and from (78) that
$$\int_{|\xi|=1} S_{n,j}(\xi)(\Delta + \omega^2) U(r\xi) \, dS = c_{n,j}\zeta_n^{(1)}(\omega r)$$
$$+ d_{n,j}\zeta_n^{(2)}(\omega r). \quad (86)$$

Since $U = o(1)$ for $r \to \infty$ we have
$$A_{n,j}(r) = o(1). \quad (87)$$

This however is only possible if $c_{n,j} = d_{n,j} = 0$. It follows therefore that
$$\int_{|\xi|=1} S_{n,j}(\xi)(\Delta + \omega^2) U(r\xi) \, dS = 0 \quad (88)$$
for all our spherical harmonics $S_{n,j}(\xi)$. Since this implies $(\Delta + \omega^2) U = 0$, we have proved Lemma 2.

We now prove

Lemma 3: Suppose the scalar function U is a solution of the differential equation
$$(\nabla + \omega^2) U = 0$$
for $|x| \ge R > 0$, which satisfies the Sommerfeld radiation condition. Then, if n represents the normal vector on the sphere $|x| = r$,
$$\nabla U = i\omega n U + O\left(\frac{1}{r^2}\right)$$
for $r \to \infty$.

On the sphere $|x| = R_1 > R$ the function U is analytic. We may therefore continue it to the inside of this sphere in such a way that the continued function is twice continuously differentiable everywhere. In $|x| \le R_1$ we form
$$\Delta U + \omega^2 U = - \rho.$$

A well-known result in the theory of the Helmholtz equation then states that
$$U(x) = \frac{1}{4\pi} \int_{|y| \le R_1} \frac{e^{i\omega|x-y|}}{|x-y|} \rho(y) \, dV_y \quad (90)$$
for all (x). For fixed (y) and $|x| \to \infty$ we have, with $(x) = r(\xi)$,
$$\frac{e^{i\omega|x-y|}}{|x-y|} = \frac{e^{i\omega r}}{r} e^{-i\omega\xi \cdot y} + O\left(\frac{1}{r^2}\right)$$
$$\nabla_x \frac{e^{i\omega|x-y|}}{|x-y|} = i\omega\xi \frac{e^{i\omega r}}{r} e^{-i\omega\xi \cdot y} + O\left(\frac{1}{r^2}\right), \quad (91)$$
where the estimates of the error terms hold uniformly with respect to all y in $|y| \le R_1$. We omit the straightforward calculation of these relations and pass directly to the asymptotic behavior of our function U, which we get by inserting (91) into (90). This then proves Lemma 3.

We are now able to show that (68), (69) follows from (70), (71). We do this in two steps. First we prove that from
$$(\nabla + \omega^2)E = 0 \quad \text{and} \quad E \cdot n = o\left(\frac{1}{r}\right) \quad (92)$$
it follows $\nabla \cdot E = 0$. Therefore (68) follows from (69) and (70). This part makes essential use of Lemma 2. Then by using Lemma 3 we deduce (6) from
$$(\Delta + \omega^2)E = 0 \quad \text{and} \quad \frac{\partial}{\partial n} E - i\omega E = o\left(\frac{1}{r}\right);$$
$$(En) = o\left(\frac{1}{r}\right). \quad (93)$$

To carry out the first step we regard the scalar function
$$U(x) = x \cdot E, \quad (94)$$
which is given by the scalar product of the vectors x and E. Then
$$\Delta U = x \cdot \Delta E + 2\nabla \cdot E = -\omega^2 x \cdot E + 2\nabla E. \quad (95)$$
Therefore U satisfies
$$(\Delta + \omega^2) U = 2\nabla \cdot E, \cdot \quad (96)$$
and we get by applying the operator $\Delta + \omega^2$ again,
$$(\Delta + \omega^2)^2 U = 0. \quad (97)$$

The asymptotic property given in (92) implies $U = o(1)$. The function U thus satisfies the conditions of Lemma 2 and we get
$$(\Delta + \omega^2) U = 0, \quad (98)$$
so that (96) gives $\nabla E = 0$.

Each of the three Cartesian co-ordinates (E_1, E_2, E_3) of E satisfies the Sommerfeld radiation condition. Lemma 3 therefore gives
$$\frac{\partial E_j}{\partial x^k} = i\omega E_j n^k + O\left(\frac{1}{r^2}\right), \quad (99)$$
which implies
$$\nabla \times E = i\omega n \times E + O\left(\frac{1}{r^2}\right) \quad (100)$$
or
$$n \times \nabla \times E = -i\omega E + o\left(\frac{1}{r}\right), \quad (101)$$

since $E \cdot n = o\left(\frac{1}{r}\right)$.

The vector radiation pattern $F(\xi)$ of the field E is given by its three Cartesian components F_1, F_2, F_3. The necessary and sufficient conditions for these scalar patterns are established in Theorem 1 which we proved previously. If we now impose the additional condition that

$$n \cdot F(\xi) = 0, \tag{102}$$

we get the full description of the vector radiation patterns as given in Theorem 2.

III. Radiation by a Finite Number of Dipoles

Suppose the dipoles j_v are located at the points (x_v) with $v = 1, 2, \cdots N$. If $\epsilon = \mu = 1$, the field E, H generated by j_v is

$$E_v = i\omega j_v \phi_v + \frac{i}{\omega}\nabla(j_v \nabla \phi_v) \qquad H_v = -j_v \times \nabla \phi_v, \tag{103}$$

with

$$\phi_v(x) = \frac{e^{i\omega|x-x_v|}}{|x-x_v|} \quad \text{and} \quad \nabla = \nabla_x. \tag{104}$$

We denote by

$$S_{v,v} = \lim_{R \to \infty} \int_{|x|=R} n(E_v \times \overline{H}_v)\, dS \tag{105}$$

the total radiation of the field E_v, H_v. From the radiation conditions

$$n \times H_v = -E_v + o\left(\frac{1}{r}\right)$$
$$\qquad\qquad\qquad\qquad\qquad r \to \infty, \tag{106}$$
$$n \times E_v = \quad H_v + o\left(\frac{1}{r}\right)$$

which hold uniformly with respect to all directions, it follows that

$$S_{v,v} = \lim_{R \to \infty} \int_{|x|=R} H_v \overline{H}_v\, dS. \tag{107}$$

Put $(x) = r(\xi)$ where (ξ) is the unit direction vector and $|x| = r$. Then for $r \to \infty$

$$\nabla \phi_v(r\xi) = i\omega\xi \frac{e^{i\omega|r\xi-x_v|}}{|r\xi - x_v|} + O\left(\frac{1}{r^2}\right) \tag{108}$$

and

$$\frac{e^{i\omega|r\xi-x_v|}}{|r\xi - x_v|} = \frac{e^{i\omega|r\xi-x_v|}}{r} + O\left(\frac{1}{r^2}\right)$$
$$= \frac{e^{i\omega r}}{r} e^{i\omega(|r\xi-x_v|-r)} + O\left(\frac{1}{r^2}\right) \tag{109}$$
$$= \frac{e^{i\omega r}}{r} e^{-i\omega(\xi x_v)} + O\left(\frac{1}{r^2}\right);$$

because

$$|r\xi - x_v| = r\left|\xi - \frac{1}{r}x_v\right| = r\sqrt{1 - \frac{2}{r}(\xi x_v) + \frac{|x_v|^2}{r^2}} \tag{110}$$

$$= r - (\xi x_v) + o\left(\frac{1}{r^2}\right).$$

We get from (103) therefore

$$H_v = i\omega(\xi \times j_v)\frac{e^{i\omega r}}{r} e^{-i\omega(\xi x_v)} + o\left(\frac{1}{r^2}\right), \tag{111}$$

and obtain from (107)

$$S_{v,v} = \frac{\omega^2}{R^2}\int_{|x|=R} |\xi \times j_v|^2\, dS. \tag{112}$$

If Ω denotes the unit sphere and $d\omega$ its surface element we have $dS = R^2\, d\omega$ and thus get

$$S_{v,v} = \omega^2 \int_\Omega |\xi \times j_v|^2\, d\omega. \tag{113}$$

Using the identity

$$|\xi \times j_v|^2 = |j_v|^2 - |(\xi j_v)|^2 \tag{114}$$

we find

$$S_{v,v} = \omega^2\left(4\pi |j_v|^2 - \int_\Omega (\overline{j_v}\xi)(j_v\xi)\, d\omega\right). \tag{115}$$

Now it is true that

$$\int_\Omega (\overline{j_v}\xi)(j_v\xi)\, d\omega = \overline{j_v} \cdot \int_\Omega \xi(j_v\xi)\, d\omega. \tag{116}$$

But

$$\int_\Omega \xi(j_v\xi)\, d\omega = \int_{|x|=1} n(j_v x)\, dS$$
$$= \int_{|x|\leq 1} \nabla(j_v x)\, dV = \int_{|x|\leq 1} j_v\, dV = \frac{4\pi}{3} j_v, \tag{117}$$

so that

$$\int_\Omega (\overline{j_v}\xi)(j_v\xi)\, d\omega = \frac{4\pi}{3} |j_v|^2. \tag{118}$$

We therefore get from (115)

$$S_{v,v} = \omega^2\frac{8\pi}{3} |j_v|^2. \tag{119}$$

This value gives the total radiation of the dipole j_v if there is no other source in the whole space. If there is another dipole j_μ located at (x_μ) the total radiation of the two dipoles regarded as coherent is in general not equal to the sum of the radiation of each of them, since these two fields may interact and so increase or decrease the total radiation.[2]

To formulate this more precisely we discuss the field $E_v + E_\mu$, $H_v + H_\mu$ which is generated by the dipoles j_v and j_μ at (x_v) and (x_μ). Let us call

$$C = \lim_{R \to \infty}\int_{|x|=R} n\left[(E_v + E_\mu)\times(\overline{H}_v + \overline{H}_\mu)\right] dS$$
$$= \lim_{R \to \infty}\int_{|x|=R} |H_v + H_\mu|^2\, dS \tag{120}$$

the total coherent radiation. Then

$$I = S_{v,v} + S_{\mu,\mu} = \frac{8\pi}{3}\omega^2(|j_v|^2 + |j_\mu|^2) \tag{121}$$

would be the total incoherent radiation.

We call

$$W = \frac{C}{I} \tag{122}$$

---

[2] If they are incoherent there is in the time average no interaction and thus the incoherent total radiation is the sum of the individual radiations of the dipoles.

the efficiency of our system. In order to calculate this quantity we have to evaluate C. From (120) it follows

$$C = \lim_{R\to\infty} \int_{|x|=R} \Big[n(E_v \times \bar{H}_v) + n(E_v \times \bar{H}_\mu)$$
$$+ n(E_\mu \times \bar{H}_v) + n(E_\mu \times \bar{H}_\mu) \Big] dS \quad (123)$$
$$= S_{v,v} + S_{\mu,\mu} + S_{v,\mu} + S_{\mu,v},$$

if we use the abbreviation

$$S_{v,\mu} = \lim_{R\to\infty} \int_{|x|=R} n(E_v \times \bar{H}_\mu)\, dS. \quad (124)$$

The radiation conditions (106) lead to

$$S_{v,\mu} = \lim_{R\to\infty} \int_{|x|=R} n(E_v \times \bar{H}_\mu)\, dS$$
$$= \lim_{R\to\infty} \int_{|x|=R} H_v \bar{H}_\mu\, dS \quad (125)$$

so that

$$S_{v,\mu} = \bar{S}_{\mu,v}. \quad (126)$$

From (111) we get

$$S_{v,\mu} = \lim_{R\to\infty} \int_{|x|=R} H_v \bar{H}_\mu\, dS$$
$$= \lim_{R\to\infty} \frac{\omega^2}{R^2} \int_{|x|=R} (\xi \times j_v)(\xi \times \bar{j}_\mu) e^{i\omega\xi(x_\mu - x_v)}\, dS$$
$$= \omega^2 \int_\Omega (\xi \times j_v)(\xi \times \bar{j}_\mu) e^{i\omega\xi(x_\mu - x_v)}\, d\omega$$
$$= \omega^2 \int_\Omega \Big[(j_v \bar{j}_\mu) - (\xi j_v)(\overline{\xi j_\mu}) \Big] e^{i\omega\xi(x_\mu - x_v)}\, d\omega. \quad (127)$$

Let

$$\Psi_n(\rho) = \sqrt{\frac{\pi}{2\rho}}\, J_{(2n+1)/2}(\rho). \quad (128)$$

Then

$$\int_\Omega e^{i\omega\xi(x_\mu - x_v)}\, d\omega = 4\pi \Psi_0(\omega\,|\,x_\mu - x_v\,|). \quad (129)$$

In order to evaluate the last integral in (127) we observe that

$$(xj_v)(x\bar{j}_\mu) - \tfrac{1}{3}\,|\,x\,|^2 (j_v \bar{j}_\mu) \quad (130)$$

is a harmonic polynomial of degree 2. Therefore

$$(\xi j_v)(\xi \bar{j}_\mu) - \tfrac{1}{3}(j_v \bar{j}_\mu) = S_2(\xi) \quad (131)$$

is a spherical harmonic of order 2, and we get

$$S_{v,\mu} = \int_\Omega [\tfrac{2}{3}(j_v \bar{j}_\mu) - S_2(\xi)] e^{i\omega\xi(x_\mu - x_v)}\, d\omega. \quad (132)$$

Using well-known formulas of the theory of Bessel functions and spherical harmonics we are led to

$$S_{v,\mu} = \frac{8\pi}{3} \omega^2 [\Psi_0(\omega\,|\,x_\mu - x_v\,|)(j_v \cdot \bar{j}_\mu)$$
$$- \frac{3}{2} S_2\left(\frac{x_\mu - x_v}{|\,x_\mu - x_v\,|}\right) \Psi_2(\omega\,|\,x_\mu - x_v\,|)]$$
$$= \frac{8\pi}{3} \omega^2 \Big[(\Psi_0(\omega\,|x_\mu - x_v\,|) - \tfrac{1}{2}\Psi_2(\omega\,|\,x_\mu - x_v\,|))(j_v \cdot \bar{j}_\mu)$$
$$+ \frac{3}{2}\,((x_\mu - x_v)j_v)((x_\mu - x_v)\bar{j}_\mu)\frac{\Psi_2(\omega\,|\,x_\mu - x_v\,|)}{|\,x_\mu - x_v\,|^2} \Big]. \quad (133)$$

This formula shows that $S_{v\mu}$ is a bilinear form which we can best describe by introducing the matrix

$$M_{v\mu} = \Big((\Psi_0(\omega R_{\mu v}) = \tfrac{1}{2}\Psi_2(\omega R_{\mu v}))\delta_{ik}$$
$$+ \frac{3}{2}\frac{\Psi_2(\omega R_{\mu v})}{R_{\mu v}{}^2}(x_\mu - x_v)_i(x_\mu - x_v)_k \Big). \quad (134)$$

We used here the abbreviations

$$R_{\mu v} = |\,x_\mu - x_v\,| \quad \text{and} \quad (x_\mu - x_v)_i = (e_i(x_\mu - x_v)), \quad (135)$$

where the vectors (e_i) denote direction vectors of the co-ordinate axes.

If $M_{v\mu}'$ denotes the transposed matrix, we get from (32)

$$M_{v\mu}' = M_{v\mu}. \quad (136)$$

If $\mu = v$ the matrix $M_{\mu v}$ becomes the unit matrix. It is therefore, possible, by using the notations of the calculus of matrices, to write the total radiation of two dipoles, j_v and j_μ, in the form

$$C = \frac{8\pi}{3} \omega^2 [(j_v, M_{vv}\bar{j}_v) + (j_v, M_{v\mu}\bar{j}_\mu)$$
$$+ (j_\mu, M_{\mu v}\bar{j}_v) + (j_\mu, M_{\mu\mu}\bar{j}_\mu)]. \quad (137)$$

The total radiation of a system of N dipoles treated at the N different points (x_v) is therefore given by

$$C = \frac{8\pi}{3} \omega^2 \sum_{v=1}^N \sum_{\mu=1}^N (j_v, M_{v\mu}\bar{j}_\mu). \quad (138)$$

This is a quadratic form which again is best described by introducing a $3N$-dimensional vector

$$J = (j_1, j_2, \cdots j_N), \quad (139)$$

and a $3N$-dimensional matrix

$$M = \begin{pmatrix} M_{11} & M_{12} & \cdots & M_{1N} \\ M_{21} & M_{22} & \cdots & M_{2N} \\ M_{N1} & M_{N2} & \cdots & M_{NN} \end{pmatrix} \quad (140)$$

Thus we get

$$C = \frac{8\pi}{3} \omega^2 (J, M\bar{J})$$
$$I = \frac{8\pi}{3} \omega^2 (J, \bar{J}) \quad (141)$$

and

$$W = \frac{(J, M\bar{J})}{(J, \bar{J})}. \quad (142)$$

The efficiency can therefore be discussed by means of the matrix M. In order to get a full description of the radiation by our dipoles we introduce the eigenvectors J_k and the eigenvalues λ_k of the matrix M. They are real and satisfy

$$J_k = \lambda_k M J_k \qquad k = 1, 2, \cdots 3N.$$
$$(J_k, J_o) = \delta_{k\sigma} \quad (143)$$

and the corresponding efficiencies are

$$W_k = \frac{1}{\lambda_k}. \quad (144)$$

By definition the W_k are never negative. On the other hand the sum of all the W_k is equal to the trace of the matrix M. As M_{vv} is the unit matrix the elements in the diagonal M are all equal to one.

We have therefore

Lemma 1: $W_1 + W_2 + \cdots + W_{3N} = 3N$.

If we assume the non-negative eigenvalues to be arranged in descending order this leads to

Lemma 2: $W_1 \geq 1$; $\qquad W_{3N} \leq 1$.

If $W_{3N} = 0$ this would mean that the field E, H generated by the dipoles corresponding to the vector J_{3N} does not radiate at all. It follows from the uniqueness theorem then that the field vanishes everywhere and thus cannot have the dipole singularities which we assumed. This is a contradiction and we get therefore $W_{3N} > 0$.

Before discussing these concepts further let us deal with the case of the two dipoles in greater detail.

We assume (x_1) to be the origin and put

$$(x_2) = Re_1 \tag{145}$$

The matrix M_{12} then has the form

$$\begin{pmatrix} A+B & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & A \end{pmatrix}$$

with

$$A = \Psi_o(\omega R) - \tfrac{1}{2}\Psi_2(\omega R) \tag{147}$$

$$B = \frac{3}{2}\Psi_2(\omega R),$$

and our matrix M becomes

$$M = \begin{pmatrix} 1 & 0 & 0 & A+B & 0 & 0 \\ 0 & 1 & 0 & 0 & A & 0 \\ 0 & 0 & 1 & 0 & 0 & A \\ A+B & 0 & 0 & 1 & 0 & 0 \\ 0 & A & 0 & 0 & 1 & 0 \\ 0 & 0 & A & 0 & 0 & 1 \end{pmatrix} \tag{148}$$

The eigenvectors and eigenvalues are readily calculated, as in Table I.

TABLE I

Eigenvector	(Eigenvalue)$^{-1}$
$(1, \quad 0, \quad 0, \quad 1, \quad 0, \quad 0)$	$1 + A + B$
$(1, \quad 0, \quad 0, -1, \quad 0, \quad 0)$	$1 - A - B$
$(0, \quad 1, \quad 0, \quad 0, \quad 1, \quad 0)$	$1 + A$
$(0, \quad 0, \quad 1, \quad 0, \quad 0, \quad 1)$	$1 + A$
$(0, \quad 1, \quad 0, \quad 0, -1, \quad 0)$	$1 - A$
$(0, \quad 0, \quad 1, \quad 0, \quad 0, -1)$	$1 - A$

Formulating this in terms of our dipoles we get the cases of Table II.

TABLE II

Arrangement of dipoles	Efficiency
$\longrightarrow \qquad \longrightarrow$	$1 + A + B$
$\longrightarrow \qquad \longleftarrow$	$1 - A - B$
$\uparrow\text{- - - - - - - - -}\uparrow$	$1 + A$
$\uparrow\text{- - - - - - - - -}\downarrow$	$1 - A$

The last two types must be considered as representing four different cases, according to whether the vectors are parallel in the plane of the paper or in a plane perpendicular to the paper.

Expressing (147) in terms of elementary functions we obtain

$$A = \frac{3}{2\omega R}\left(\sin \omega R + \frac{\cos \omega R}{\omega R} - \frac{\sin \omega R}{(\omega R)^2}\right)$$

$$= 1 - \frac{7}{10}(\omega R)^2 + v$$

$$A + B = \frac{3}{(\omega R)^2}\left(\frac{\sin \omega R}{\omega R} - \cos \omega R\right)$$

$$= 1 - \frac{1}{10}(\omega R)^2 + \cdots . \tag{149}$$

The signs of these expressions will determine which of the four cases mentioned above is most efficient. As it is impossible that both of the expressions (149) vanish simultaneously, there is always at least one arrangement of dipoles which has an efficiency W with $W > 1$.

Each of the eigenvectors J_k represents an arrangement of dipoles which generates a certain pattern $F_k(\xi)$. On the other hand, any configuration of two dipoles can be written in the form $\Sigma a^k J_k$ $(k = 1, 2, \cdots 6)$ where the a^k are suitably chosen complex constants. The corresponding pattern would then be $\Sigma a^k F_k(\xi)$ and consequently

$$C = \int_{|\xi|=1} \Big| \sum a^k F_k(\xi)\Big|^2 \, dS = \sum_k \sum_\lambda a^k a^\lambda A_{k\lambda} \tag{150}$$

with

$$A_{k\lambda} = \int_{|\xi|=1} F_k(\xi)\overline{F_\lambda(\xi)} \, dS. \tag{151}$$

From (141) and (143) we get

$$C = \frac{8\pi\omega^2}{3}\left(\sum a^k J_k, \sum C\overline{a}^\lambda \overline{J}_\mu\right) = \frac{8\pi}{3}\omega^2\sum_{k=1}^{6}\lambda_k|a^\lambda|^2. \tag{152}$$

As this holds for all a^λ we obtain by comparing (145) and (152)

$$A_{k\mu} = \frac{8\pi}{3}\omega^2\lambda_k\delta_{k\lambda} . \tag{153}$$

In particular this gives

$$\int_{|\xi|=1} F_k(\xi)\overline{F_\lambda(\xi)} \, dS = 0 \quad \text{for } k \neq \lambda. \tag{154}$$

The radiation patterns $F_k(\xi)$ may thus be regarded as an orthogonal basis for all the patterns which can be generated by two dipoles at fixed points.

It is now obvious how the more general case of N dipoles has to be treated and we only mention that the $3N$ patterns corresponding to the eigenvectors J_k again represent an orthogonal basis for all those patterns which can be generated by N dipoles at fixed points.

A Refinement of the WKB Method and Its Application to the Electromagnetic Wave Theory

ISAO IMAI†

Summary—When dealing with the problem of diffraction of waves, certain special functions appear which are defined by ordinary differential equations of the second order; for example, Bessel functions, Legendre functions, and Mathieu functions appear for the case of a circular cylinder, a sphere, and an elliptic cylinder, respectively. Exact solutions are obtained in the form of infinite series of such functions, which are, in general, poorly convergent when the wavelength is comparable with or smaller than the dimension of the body. In such a case the series can be transformed into contour integrals and then evaluated by the method of steepest descents or by forming the residue series. For this purpose asymptotic expressions for the special functions are needed. A similar situation arises also in the problem of propagation of short radio waves in a horizontally stratified atmosphere.

In this paper a refinement of the WKB method is presented which enables one to obtain very accurate and compact expressions for such functions, which are particularly suited for the evaluation of zeros. The application of the method is illustrated for the case of Bessel functions, parabolic cylinder functions, Coulomb wave functions, etc.

1. Introduction

I T IS well known that the diffraction of electromagnetic waves by a body of simple shape can often be exactly dealt with by the method of separation of variables. Thus, introducing appropriate curvilinear co-ordinates, we can obtain the solution in the form

$$\psi = \sum_m C(m) X_m(\xi) Y_m(\eta) \tag{1}$$

or in the form of a series of such functions. Here X_m and Y_m satisfy certain differential equations of the second order of the form

$$\frac{d^2\Phi}{dx^2} + a^2 P(x)\Phi = 0, \tag{2}$$

where a is a parameter related to the constant m. Also, $C(m)$ is a constant with respect to ξ and η but depending on m. Various special functions appear for different bodies. Thus, Bessel functions appear for a circular cylinder, Legendre functions for a sphere, Mathieu functions for an elliptic cylinder, etc.

For the case in which the wavelength λ is large in comparison with the dimension of the diffracting body, the infinite series (1) is known to be rapidly convergent. But, if the wavelength is comparable with or less than the dimension of the body, the series is poorly convergent and the numerical computation requires a prohibitive amount of labor. Indeed, tabulation of special functions to very high order m is a prerequisite for such numerical work. As an example of this kind of research, the famous beautiful work of Meixner and Andrejewski[1] on the diffraction by a disk may be mentioned.

The above-mentioned difficulty may be overcome by a technique named after Watson. Thus, the infinite series is first transformed into a contour integral in a complex plane:

$$\psi = \frac{1}{2i} \int C(\alpha) X_\alpha(\xi) Y_\alpha(\eta) \frac{i^{-2\alpha} \, d\alpha}{\sin \pi\alpha} \tag{3}$$

and the path of integration is then suitably transformed so that (3) may be evaluated by the method of steepest descents or by the theorem of residues. But, for carrying out this procedure, it is essential to have the knowledge of the properties of X_α, Y_α, $C(\alpha)$ as functions of α, in particular their asymptotic behavior as $\alpha \to \infty$, as well as the zeros of these functions. This procedure has actually been carried out for a circular cylinder by Franz[2] and the present author,[3] and for a sphere by Franz.[2]

It may be noted that Franz's and the present author's treatment has some essential difference from earlier investigations of similar nature in that some part of the residue series is again converted into a contour integral so that the remaining part may be convergent even in the illuminated region, whereas in the earlier investigations residue series was convergent only in the shadow region.

Although the above-mentioned method seems to be very promising for the case of short wavelength, it cannot be employed for a circular disk or an elliptic cylinder, since sufficient knowledge concerning the appropriate special functions is not available at present. Therefore, it would be desirable to have some technique for investigating the asymptotic behavior of such functions as defined by (1).

In the problem of propagation of short radio waves in a horizontally stratified atmosphere, we meet with equations of the type[4]

$$\frac{d^2 v}{dz^2} + (k^2 N^2 - \kappa^2)v = 0. \tag{4}$$

In fact, the electromagnetic field generated by a dipole antenna near the earth can be determined from a Hertzian potential ψ expressed in the form

$$\psi = \int_0^\infty \kappa J_0(\kappa r) v(z, \kappa) \, d\kappa, \tag{5}$$

† Univ. of Tokyo, Tokyo, Japan.

[1] J. Meixner and W. Andrejewski, "Strenge Theorie der Beugung ebener elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Oeffnung in vollkommen leitenden ebenen Schirm," *Ann. Phys.*, vol. 7, pp. 157–168; April, 1950. W. Andrejewski, "Die Beugung elektromagnetischer Wellen an der leitenden Kreisscheibe und an der kreisförmigen Oeffnung in leitenden ebenen Schirm," *Z. angew. Phys.*, vol. 5, pp. 178–186; May, 1953.

[2] W. Franz, "Ueber die Greenschen Funktionen der Zylinder und der Kugel," *Z. Naturf.*, vol. 9a, pp. 705–716; September, 1954.
[3] I. Imai, "Die Beugung elektromagnetischer Wellen an einem Kreiszylinder," *Zeit. für Phys.*, vol. 137, pp. 31–48; February, 1954.
[4] D. E. Kerr, "Propagation of Short Radio Waves," M.I.T. Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, N. Y., p. 58 ff; 1951. C. L. Pekeris, "Asymptotic solutions for the normal mode in the theory of microwave propagation," *Jour. Appl. Phys.*, vol. 17, pp. 1108–1124; December, 1946.

where r and z are respectively the horizontal and vertical distances from the dipole, and $v(z, \kappa)$ is the solution of (4) satisfying appropriate boundary conditions. Here N is the modified index of refraction and $k = 2\pi/\lambda$, λ being the wavelength. It may be easily seen that here also we are concerned with equations of the form (1).

The object of this paper is to present an account of a refinement of the WKB method recently developed by the author.[5] The main features of the method will be presented in a form convenient for application, and the procedure will be illustrated by choosing examples from Bessel functions, parabolic cylinder functions and Coulomb wave functions. The resulting formulas are believed to be new.

The method enables one to obtain solutions of (1) that are exact to the order of $0(a^{-3})$ compared with unity. Therefore, it may be regarded as a presentation in a compact form of higher-order approximations of Langer's method.[6]

2. Outline of an Improved Approximation of the WKB Method

Let us consider the second-order linear differential equation

$$\frac{d^2\Phi}{dx^2} + a^2 P(x)\Phi = 0, \qquad (6)$$

where a is a parameter and $P(x)$ has the form

$$P(x) = a_1 x + a_2 x^2 + a_3 x^3 + \cdots$$
$$= a_1 x(1 + b_1 x + b_2 x^2 + \cdots). \qquad (7)$$

If we introduce the variables z and Ψ defined by

$$z = \int_0^x P^{1/2}\, dx, \qquad (8)$$

$$\Phi = P^{-1/4}\, \Psi, \qquad (9)$$

(6) becomes

$$\frac{d^2\Psi}{dz^2} + (a^2 - Q)\Psi = 0, \qquad (10)$$

where

$$Q = P^{-1/4}\frac{d^2 P^{1/4}}{dz^2} = -P^{-3/4}\frac{d^2 P^{-1/4}}{dx^2}. \qquad (11)$$

Inserting (7) in (8), we have

$$z = \frac{2}{3} a_1^{1/2} x^{3/2}\Big\{1 + \frac{3}{10} b_1 x + \Big(\frac{3}{14} b_2 - \frac{3}{56} b_1^2\Big)x^2$$

$$+ \Big(\frac{1}{6} b_3 - \frac{1}{12} b_1 b_2 + \frac{1}{48} b_1^3\Big)x^3 + \Big(\frac{3}{22} b_4 - \frac{3}{88} b_2^2 \quad (12)$$

$$- \frac{3}{44} b_1 b_3 + \frac{9}{176} b_1^2 b_2 - \frac{15}{1408} b_1^4\Big)x^4 + \cdots\Big\},$$

[5] I. Imai, "On a refinement of the WKB method," *Phys. Rev.*, vol. 74, pp. 113; July 1, 1948. I. Imai, "Asymptotic solutions of ordinary linear differential equations of the second order," *Phys. Rev.*, vol. 80, pp. 1112; December 15, 1950.

[6] R. E. Langer, "On the asymptotic solutions of ordinary differential equations with application to Bessel functions of large order," *Trans. Amer. Math. Soc.*, vol. 33, pp. 23–64; January, 1931.

which can be solved for x to give

$$x = t - \frac{1}{5} b_1 t^2 - \Big(\frac{1}{7} b_2 - \frac{22}{175} b_1^2\Big) t^3$$

$$- \Big(\frac{1}{9} b_3 - \frac{67}{315} b_1 b_2 + \frac{823}{7875} b_1^3\Big) t^4$$

$$- \Big(\frac{1}{11} b_4 - \frac{48}{539} b_2^2 - \frac{94}{495} b_1 b_3$$

$$+ \frac{34933}{121275} b_1^2 b_2 - \frac{301306}{3031875} b_1^4\Big) t^5 + \cdots, \qquad (13)$$

with

$$t = \Big(\frac{3}{2} a_1^{-1/2} z\Big)^{2/3}. \qquad (14)$$

Substituting (7) and (13) in (11), we find that

$$Q = -\Big(\frac{5}{36}\frac{1}{z^2} + \lambda z^{-2/3} + \lambda_0 + \lambda_1 z^{2/3} + \cdots\Big), \qquad (15)$$

where

$$\lambda = -\frac{3}{35}\Big(\frac{3}{2} a_1\Big)^{-2/3}(5b_2 - 3b_1^2), \qquad (16)$$

$$\lambda_0 = -\frac{4}{75}\frac{1}{a_1}(25b_3 - 35b_1 b_2 + 14b_1^3), \qquad (17)$$

$$\lambda_1 = -\frac{27}{26950}\Big(\frac{3}{2} a_1\Big)^{-4/3}(6125b_4 - 4350b_2^2$$

$$- 9800b_1 b_3 + 12080b_1^2 b_2 - 3624b_1^4). \qquad (18)$$

It can be shown that for $a \to \infty$ (10) has two independent solutions Ψ_1 and Ψ_2 such that

$$\Psi_{1,2} = \exp\Big[\pm iaz \pm \frac{1}{2ia} Q_1(z) + \frac{1}{4a^2} Q(z)\Big] + 0(a^{-3})$$

$$\text{for} \quad z \neq 0; \qquad (19)$$

$$\Psi_{1,2} = \Big(\frac{\pi a}{2}\Big)^{1/2} i^{\pm 5/6} z^{1/6} \zeta H_{1/3}^{(1,2)}(\eta) + 0(a^{-3})$$

$$\text{for} \quad z = 0(a^{-1}). \qquad (20)$$

Here

$$Q_1 = \int_0^z \Big(Q + \frac{5}{36}\frac{1}{z^2}\Big) dz + \frac{5}{36}\frac{1}{z}, \qquad (21)$$

$$\zeta = \Big(\xi - \frac{1}{5}\lambda_1 \kappa^{-2}\xi^2\Big)^{1/2}, \qquad (22)$$

$$\eta = \kappa\Big(\xi + \frac{1}{5}\lambda_1 \kappa^{-2}\xi^2\Big)^{3/2}, \qquad (23)$$

$$\xi = z^{2/3} + \lambda\kappa^{-2}, \qquad (24)$$

$$\kappa^2 = a^2 + \lambda_0. \qquad (25)$$

It is to be noted that in order that (20) may smoothly go over into (19) we must require that $|\arg \eta| < \pi$ and hence $|\arg az| < \pi$. If η does not satisfy this condition we have only to make use of the relations

$$H_\nu^{(1)}(e^{m\pi i}z) = -\frac{\sin(m-1)\nu\pi}{\sin\nu\pi} H_\nu^{(1)}(z)$$

$$- e^{-\nu\pi i}\frac{\sin m\nu\pi}{\sin\nu\pi} H_\nu^{(2)}(z),$$

$$H_\nu^{(2)}(e^{m\pi i}z) = e^{\nu\pi i}\frac{\sin m\nu\pi}{\sin\nu\pi} H_\nu^{(1)}(z)$$

$$+ \frac{\sin(m+1)\nu\pi}{\sin\nu\pi} H_\nu^{(2)}(z),$$

and apply the formulas (19) and (20) to the function $H_{1/3}^{(1,2)}(e^{-m\pi i}\eta)$ where $|\arg(e^{-m\pi i}\eta)| < \pi$.

When considering the case $x \to 0$, so that $z \to 0$, it is convenient to use $J_{\pm 1/3}(\eta)$ instead of $H_{1/3}^{(1,2)}(\eta)$ in (20). Thus, we have for $x \to 0$

$$P^{-1/4}z^{1/6}\zeta J_{1/3}(\eta) = (\tfrac{3}{2}a_1)^{-1/6}\frac{2^{-1/3}}{\Gamma(\tfrac{4}{3})}\lambda\kappa^{-5/3}\big[1 - \{\tfrac{1}{5}b_1$$

$$-\tfrac{2}{3}\lambda^{-1}(\tfrac{3}{2}a_1)^{1/3}\kappa^2\}x + \cdots\big], \quad (26)$$

$$P^{-1/4}z^{1/6}\zeta J_{-1/3}(\eta)$$

$$= (\tfrac{3}{2}a_1)^{-1/6}\frac{2^{1/3}}{\Gamma(\tfrac{2}{3})}\kappa^{-1/3}\big[1 - \{\tfrac{1}{5}b_1$$

$$+ (\tfrac{3}{4}\lambda^2 + \tfrac{2}{15}\lambda_1)(\tfrac{3}{2}a_1)^{1/3}\kappa^{-2}\}x + \cdots\big]. \quad (27)$$

The behavior as $x \to 0$ of the solutions Φ_1 and Φ_2 can be deduced from (26) and (27) by use of the relations

$$H_{1/3}^{(1)} = \frac{2}{\sqrt{3}}i^{2/3}(i^{-1/3}J_{1/3} - i^{1/3}J_{-1/3}), \quad (28)$$

$$H_{1/3}^{(2)} = \frac{2}{\sqrt{3}}i^{-2/3}(i^{1/3}J_{1/3} - i^{-1/3}J_{-1/3}). \quad (29)$$

The general solution of (6) can be expressed in the form

$$\Phi(x) = A_1\Phi_1(x) + A_2\Phi_2(x)$$
$$= P^{-1/4}e^{Q/4a^2}\{A_1e^{i(az-Q_1/2a)} + A_2e^{-i(az-Q_1/2a)}\}$$
$$\text{for } x \neq 0, \quad (30)$$

$$\Phi(x) = \left(\frac{\pi a}{2}\right)^{1/2}P^{-1/4}z^{1/6}\zeta \{A_1i^{5/6}H_{1/3}^{(1)}(\eta)$$

$$+ A_2i^{-5/6}H_{1/3}^{(2)}(\eta)\} \quad \text{for } x = 0(a^{-2/3}). \quad (31)$$

Hence the first zeros x_s $(s = 1, 2, \cdots)$ of $\Phi(x)$ can be calculated from the roots η_s of

$$A_1i^{5/6}H_{1/3}^{(1)}(\eta) + A_2i^{-5/6}H_{1/3}^{(2)}(\eta) = 0 \quad (32)$$

through (23), (24), (13) and (14). For this purpose it may be convenient to use the relation

$$t = \left(\frac{3}{2}a_1^{-1/2}\right)^{2/3}(\eta^{2/3}a^{-2/3} - \lambda a^{-2} - \tfrac{1}{3}\lambda_0\eta^{2/3}a^{-8/3}$$

$$-\tfrac{1}{5}\lambda_1\eta^{4/3}a^{-10/3} + \cdots), \quad (33)$$

which can be derived from (14), (23), (24), and (25). Since the coefficients of t^n in (13) and those of $a^{-2n/3}$ in (33) are all pure numbers, it is very easy to express x_s as a power series in $a^{-2/3}$ in the form

$$x_s = c_1 a^{-2/3} + c_2 a^{-4/3} + \cdots, \quad (34)$$

where the numerical values of $c_1, c_2, \cdots, c_5$ can be explicitly determined from the constants a_1, b_1, b_2, b_3, b_4.

Further, it may be remarked that larger zeros of $\Phi(x)$ can be found by using (30).

3. Bessel Functions

Bessel functions of order a satisfy (6) with

$$\Phi(x) = Z_a(ae^x), \quad (35)$$

and

$$P(x) = e^{2x} - 1. \quad (36)$$

Comparing (36) with (7) we have at once

$$a_1 = 2, b_1 = 1, b_2 = \frac{2}{3}, b_3 = \frac{1}{3}, b_4 = \frac{2}{15}. \quad (37)$$

Hence (16), (17), and (18) give

$$\lambda = -\frac{3^{1/3}}{105}, \lambda_0 = \frac{2}{75}, \lambda_1 = -\frac{69 \times 3^{2/3}}{13475}. \quad (38)$$

It is convenient to put

$$e^x = \sec\theta. \quad (39)$$

Then (36), (8), (11), and (21) become respectively

$$P = \tan^2\theta, \quad (40)$$

$$z = \tan\theta - \theta, \quad (41)$$

$$Q = -\frac{\cos^2\theta(\cos^2\theta + \tfrac{1}{4})}{\sin^6\theta}, \quad (42)$$

$$Q_1 = \frac{5}{12}\cot^3\theta + \frac{1}{4}\cot\theta. \quad (43)$$

Now the well-known asymptotic expansions of Hankel functions give

$$H_a^{(1,2)}(a\sec\theta) \sim \left(\frac{\pi a\sec\theta}{2}\right)^{-1/2}\exp\left[\mp\frac{\pi}{2}i\left(a + \frac{1}{2}\right)\right.$$

$$\left.\pm ia\sec\theta\right] \sim \left(\frac{\pi a}{2}\right)^{-1/2}P^{-1/4}\exp\left(\pm iaz \mp \frac{\pi}{4}i\right)$$

as $\theta \to \pi/2$. Comparing this with (19), we can easily find that

$$H_a^{(1,2)}(a\sec\theta) = \left(\frac{2}{\pi a}\right)^{1/2}\cot^{1/2}\theta\exp\left[\pm iaz \mp \frac{\pi}{4}i\right.$$

$$\left.\mp\frac{i}{2a}Q_1 + \frac{1}{4a^2}Q\right] + 0(a^{-7/2}) \quad \text{for } \theta \neq 0 \quad (44)$$

and hence, by (20),

$$H_a^{(1,2)}(a\sec\theta) = i^{\pm 1/3}\cot^{1/2}\theta z^{1/6}\zeta H_{1/3}^{(1,2)}(\eta)$$

$$+ 0(a^{-7/2}) \quad \text{for } \theta = 0(a^{-1/3}). \quad (45)$$

Accordingly, for a function defined by

$$Z_a(x) = c_1H_a^{(1)}(x) + c_2H_a^{(2)}(x), \quad (46)$$

we have

$$Z_a(a\sec\theta) = \left(\frac{2}{\pi a}\right)^{1/2}\cot^{1/2}\theta\, e^{-Q/4a^2}(c_1e^{iR} + c_2e^{-iR})$$

$$\text{for } \theta \neq 0; \quad (47)$$

$$Z_a(a\sec\theta) = \cot^{1/2}\theta\, z^{1/6}\zeta \{c_1i^{1/3}H_{1/3}^{(1)}(\eta)$$

$$+ c_2i^{-1/3}H_{1/3}^{(2)}(\eta)\} \quad \text{for } \theta = 0(a^{-1/3}), \quad (48)$$

where Q is given by (42) and

$$R = a(\tan\theta - \theta) - \frac{\pi}{4} - \frac{1}{2a}\left(\frac{5}{12}\cot^3\theta + \frac{1}{4}\cot\theta\right). \quad (49)$$

In particular, if we put $c_1 = c_2 = 1/2$, we have

$$J_a(a \sec \theta) = \left(\frac{2}{\pi a}\right)^{1/2} \cot^{1/2} \theta \; e^{-Q/4a^2} \cos R$$

$$\text{for } \theta \neq 0, \quad (50)$$

$$J_a(a \sec \theta) = \frac{1}{\sqrt{3}} \cot^{1/2}\theta \; z^{1/6}\zeta \; \{J_{1/3}(\eta) + J_{-1/3}(\eta)\}$$

$$\text{for } \theta = 0(a^{-1/3}). \quad (51)$$

Putting $c_1 = -c_2 = -i/2$, we have

$$Y_a(a \sec \theta) = \left(\frac{2}{\pi a}\right)^{1/2} \cot^{1/2}\theta \; e^{Q/4a^2} \sin R$$

$$\text{for } \theta \neq 0, \quad (52)$$

$$Y_a(a \sec \theta) = \cot^{1/2}\theta \; z^{1/6} \{J_{1/3}(\eta) - J_{-1/3}(\eta)\}$$

$$\text{for } \theta = 0(a^{-1/3}). \quad (53)$$

Next let us consider the first zeros of $Z_a(a \sec \theta)$. Remembering (32), we denote the roots of

$$c_1 i^{1/3} H_{1/3}^{(1)}(\eta) + c_2 i^{-1/3} H_{1/3}^{(2)}(\eta) = 0 \quad (54)$$

by $\eta_s(s = 1, 2, \cdots)$. Then, by (13) and (33), the zeros $a \sec \theta_s$ are given explicitly in the form

$$\sec \theta_s = 1 + \frac{1}{2} (3\eta_s)^{2/3} a^{-2/3} + \frac{3}{40} (3\eta_s)^{4/3} a^{-4/3}$$

$$+ \left\{\frac{1}{70} - \frac{1}{2800} (3\eta_s)^2\right\} a^{-2}$$

$$- \left\{\frac{1}{6300} (3\eta_s)^{2/3} + \frac{479}{1008000} (3\eta_s)^{8/3}\right\} a^{-8/3}$$

$$- \left\{\frac{551}{646800} (3\eta_s)^{4/3} - \frac{20231}{258720000} (3\eta_s)^{10/3}\right\} a^{-10/3}$$

$$+ \cdots . \quad (55)$$

This is in agreement with the result obtained by Olver[7] by means of an entirely different method.

In particular, putting $\eta_1 = 2.3834466$ [the first zero of $J_{1/3}(\eta) + J_{-1/3}(\eta)$] we find the expression for the first zero of $J_a(ax)$ as function of a,

$$x_1 = 1 + 1.8557571 \, a^{-2/3} + 1.0331503 \, a^{-4/3}$$

$$- 0.0039741 \, a^{-2} - 0.090763 \, a^{-8/3}$$

$$+ 0.043339 \, a^{-10/3} + \cdots . \quad (56)$$

It seems that corresponding formula given in Jahnke-Emde[8] contains some numerical error.

Finally, application of (26) and (27) to (51) and (53) gives

$$\sqrt{3} \left.\begin{matrix} J_a(a) \\ Y_a(a) \end{matrix}\right\} = \pm \frac{2^{1/3}3^{-1/6}}{\Gamma(\frac{2}{3})} \, x^{-1/3} - \frac{2^{-1/3}3^{1/6}}{\Gamma(\frac{4}{3})} \, x^{-5/3} \quad (57)$$

and

$$\sqrt{3} \left.\begin{matrix} J_a'(a) \\ Y_a'(a) \end{matrix}\right\} = a^{-1} \left\{ \frac{2^{2/3}3^{-5/6}}{\Gamma(\frac{4}{3})} \, x^{1/3}\left(1 + \frac{1}{350} \, x^{-2}\right) \right.$$

$$\left. \mp \frac{2^{1/3}3^{-1/6}}{5\Gamma(\frac{2}{3})} \, x^{-1/3}\left(1 - \frac{71}{7700} \, x^{-2}\right)\right\}. \quad (58)$$

Further, remembering that

$$\varkappa^2 = a^2 + \lambda_0 = a^2 + \frac{2}{75},$$

we can express (57) and (58) as

$$\sqrt{3} \left.\begin{matrix} J_a(a) \\ Y_a(a) \end{matrix}\right\} = \pm \frac{2^{1/3}3^{-1/6}}{\Gamma(\frac{2}{3})} \, a^{-1/3}\left(1 - \frac{1}{225} \, a^{-2}\right)$$

$$- \frac{2^{-1/3}3^{1/6}}{\Gamma(\frac{4}{3})} \, a^{-5/3} + 0(a^{-11/3}), \quad (59)$$

$$\sqrt{3} \left.\begin{matrix} J_a'(a) \\ Y_a'(a) \end{matrix}\right\} = \frac{2^{2/3}3^{-5/6}}{\Gamma(\frac{4}{3})} \, a^{-2/3}\left(1 + \frac{23}{3150} \, a^{-2}\right)$$

$$\mp \frac{2^{1/3}3^{-1/6}}{5\Gamma(\frac{2}{3})} \, a^{-4/3}\left(1 - \frac{947}{69300} \, a^{-2}\right). \quad (60)$$

These agree with the formulas given in Jahnke-Emde[9].

4. Parabolic Cylinder Functions

Recently Rice[10] has studied the diffraction of plane radio waves by a parabolic cylinder with a view to investigating the nature of shadows that are cast by hills in microwave propagation. In this work knowledge of the behavior of parabolic cylinder functions of large complex order was required, which he studied by the method of steepest descents.

Here we shall derive the formulas by a straightforward application of the general formulas given in § 2.

Rice considers the three functions

$$U_n(w) = 2^{n/2}e^{w^2/2}D_n(2^{1/2}w), \quad (61)$$

$$V_n(w) = -\frac{i^{-n}2^{n/2}}{(2\pi)^{1/2}} \, e^{w^2/2}D_{-n-1}(-2^{1/2}iw), \quad (62)$$

$$W_n(w) = -\frac{i^n 2^{n/2}}{(2\pi)^{1/2}} \, e^{w^2/2}D_{-n-1}(2^{1/2}iw), \quad (63)$$

$$U_n(w) + V_n(w) + W_n(w) = 0, \quad (64)$$

where $D_n(z)$ is the usual parabolic cylinder function. It is known that if $T_n(w)$ denotes any one of $U_n(w)$, $V_n(w)$, and $W_n(w)$, then

$$\frac{d^2}{dw^2} \{e^{-w^2/2}T_n(w)\} + (2n + 1 - w^2)e^{-w^2/2}T_n(w) = 0.$$

$$(65)$$

It will readily be seen that if we put

$$w = (2n + 1)^{1/2}(1 - x), \quad 2n + 1 = a, \quad (66)$$

$$\Phi(x) = e^{-w^2/2}T_n(w), \quad (67)$$

$$P(x) = 2x - x^2, \quad (68)$$

(65) reduces to the normal form (6).

For convenience, we put

$$x = 1 - \cos \theta. \quad (69)$$

Then, after elementary calculation, we find that

$$z = \tfrac{1}{4}(2\theta - \sin 2\theta), \quad (70)$$

$$Q = -\frac{3\cos^2 \theta + 2}{4\sin^6 \theta}, \quad (71)$$

---

[7] F. W. J. Olver, "A further method for the evaluation of zeros of Bessel functions and some new asymptotic expansions for zeros of functions of large order," *Proc. Camb. Phil. Soc.*, vol. 47, pp. 699–712; October, 1951.

[8] E. Jahnke and F. Emde, "Funktionentafeln mit Formeln und Kurven," Teubner, Leipzig, Germany, p. 211; 1933.

[9] *Ibid.*, p. 209.

[10] S. O. Rice, "Diffraction of plane radio waves by a parabolic cylinder. Calculation of shadows behind hills," *Bell Sys. Tech. Jour.*, vol. 33, p. 417–504; March, 1954.

$$Q_1 = \frac{5}{12}\frac{\cos\theta}{\sin^3\theta} + \frac{1}{12}\cot\theta. \tag{72}$$

It is known that $U_n(w)$, $V_n(w)$, $W_n(w)$ are one-valued analytic functions of w and n, and that

$$U_n(0) = \frac{\cos\frac{n\pi}{2}}{\Gamma\left(\frac{n}{2}+1\right)}, \qquad U_n'(0) = \frac{2\sin\frac{n\pi}{2}}{\Gamma\left(\frac{n+1}{2}\right)}, \tag{73}$$

$$V_n(0) = \frac{-i^{-n}}{2\Gamma\left(\frac{n}{2}+1\right)}, \qquad V_n'(0) = \frac{-i^{-n+1}}{\Gamma\left(\frac{n+1}{2}\right)}, \tag{74}$$

$$W_n(0) = \frac{-i^n}{2\Gamma\left(\frac{n}{2}+1\right)}, \qquad W_n'(0) = \frac{-i^{n-1}}{\Gamma\left(\frac{n+1}{2}\right)}. \tag{75}$$

But, from the general formula (19), it is seen that (65) has particular solutions

$$T_{1,2} = \frac{1}{\sqrt{\sin\theta}}\exp\left[\frac{1}{4}(2n+1)(1+\cos 2\theta)\right.$$

$$\pm\frac{i}{4}(2n+1)(2\theta-\sin 2\theta) \mp \frac{i}{2(2n+1)}Q_1$$

$$\left. + \frac{1}{4(2n+1)^2}Q\right] \quad \text{for } \theta \ne 0. \tag{76}$$

Therefore $V_n(w)$ can be expressed as

$$V_n(w) = AT_1 + BT_2, \tag{77}$$

where A and B are constants to be determined as

$$A = \frac{V_n(0)T_2'(0) - V_n'(0)T_2(0)}{T_1(0)T_2'(0) - T_1'(0)T_2(0)}, \tag{78}$$

$$B = \frac{V_n'(0)T_1(0) - V_n(0)T_1'(0)}{T_1(0)T_2'(0) - T_1'(0)T_2(0)}. \tag{79}$$

Taking account of the fact that $w \to 0$ as $\theta \to \pi/2$, we can find from (77) that

$$T_{1,2}(0) = \exp\left[-\frac{1}{8(2n+1)^2} \pm \frac{\pi}{4}(2n+1)\,i\right], \tag{80}$$

$$T_{1,2}'(0) = \mp\, i\frac{4(2n+1)^2+1}{4(2n+1)^{3/2}}T_{1,2}(0). \tag{81}$$

Substituting (74), (80), (81) into (78) and (79) we have

$$A = \frac{V_n(0)}{2T_1(0)}(1-\epsilon), \quad B = \frac{V_n(0)}{2T_2(0)}(1+\epsilon), \tag{82}$$

where

$$\epsilon = \frac{8(2n+1)^{3/2}}{4(2n+1)^2+1}\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)}. \tag{83}$$

Hence (77) becomes

$$V_n(w) = -\frac{i^{-n}e^X(\cos Y - i\epsilon\sin Y)}{2\Gamma\left(\frac{n}{2}+1\right)\sqrt{\sin\theta}}, \tag{84}$$

where

$$X = \tfrac{1}{4}(2n+1)(1+\cos 2\theta) + \frac{1}{4(2n+1)^2}(Q+\tfrac12), \tag{85}$$

$$Y = \tfrac{1}{4}(2n+1)(2\theta-\pi-\sin 2\theta) - \frac{1}{2(2n+1)}Q_1. \tag{86}$$

In a similar manner it is found that

$$W_n(w) = -\frac{i^n e^X(\cos Y \dotplus i\epsilon\sin Y)}{2\Gamma\left(\frac{n}{2}+1\right)\sqrt{\sin\theta}}, \tag{87}$$

$$U_n(w) = \frac{e^X\left(\cos\frac{n\pi}{2}\cos Y - \epsilon\sin\frac{n\pi}{2}\sin Y\right)}{\Gamma\left(\frac{n}{2}+1\right)\sqrt{\sin\theta}}. \tag{88}$$

Of course ϵ can be given simpler form by use of Stirling's formula, but special caution is needed concerning the range of $|\arg n|$. Thus, we have for $|\arg n| < \pi$

$$\frac{8(2n+1)^{3/2}}{4(2n+1)^2+1}\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)} = 1 + 0\left(\frac{1}{n^3}\right). \tag{89}$$

For $0 < \arg n < 2\pi$ we may use the relation

$$\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)} = -\cot\frac{n\pi}{2}\frac{\Gamma\left(\frac{1-n}{2}\right)}{\Gamma\left(-\frac{n}{2}\right)} \tag{90}$$

together with the formula (89) with n replaced by $ne^{-\pi i}$ to obtain

$$\frac{8(2n+1)^{3/2}}{4(2n+1)^2+1}\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

$$= i\cot\frac{n\pi}{2}\left\{1 + 0\left(\frac{1}{n^3}\right)\right\}. \tag{91}$$

Thus, finally

$$\epsilon = \begin{cases} 1 & -\pi < \arg n < \pi \\ i\cot\frac{\pi}{2}n & 0 < \arg n < 2\pi. \end{cases} \tag{92}$$

The value of ϵ for other ranges of $\arg n$ may be obtained in quite a similar way.

Summarizing we have, for $\theta \ne 0$,

$$U_n(w) = \begin{cases} \dfrac{e^X \cos\left(\dfrac{n}{2}\pi + Y\right)}{\Gamma\left(\dfrac{n}{2}+1\right)\sqrt{\sin\theta}} & -\pi < \arg n < \pi \\[2em] \dfrac{\cos\dfrac{n\pi}{2}\, e^{X-iY}}{\Gamma\left(\dfrac{n}{2}+1\right)\sqrt{\sin\theta}} & 0 < \arg n < 2\pi \end{cases} \quad (93)$$

$$V_n(w) = \begin{cases} -\dfrac{i^{-n}e^{X-iY}}{2\Gamma\left(\dfrac{n}{2}+1\right)\sqrt{\sin\theta}} & -\pi < \arg n < \pi \\[2em] -\dfrac{i^{-n}e^X \sin\left(\dfrac{n}{2}\pi + Y\right)}{2\Gamma\left(\dfrac{n}{2}+1\right)\sin\dfrac{n\pi}{2}\sqrt{\sin\theta}} & \\[1em] & 0 < \arg n < 2\pi \end{cases} \quad (94)$$

$$W_n(w) = \begin{cases} -\dfrac{i^{n}e^{X+iY}}{2\Gamma\left(\dfrac{n}{2}+1\right)\sqrt{\sin\theta}} & -\pi < \arg n < \pi \\[2em] -\dfrac{i^{n}e^X \sin\left(\dfrac{n}{2}\pi - Y\right)}{2\Gamma\left(\dfrac{n}{2}+1\right)\sin\dfrac{n\pi}{2}\sqrt{\sin\theta}} & \\[1em] & 0 < \arg n < 2\pi. \end{cases} \quad (95)$$

Next, we consider the case $\theta \doteq 0$, that is, $w \doteq (2n+1)^{1/2}$. From (68) we immediately find that

$$a_1 = 2, \quad b_1 = -\tfrac{1}{2}, \quad b_2 = b_3 = \cdots = 0. \quad (96)$$

Hence, by (16), (17), and (18),

$$\lambda = \frac{3^{4/3}}{140}, \quad \lambda_0 = -\frac{7}{150}, \quad \lambda_1 = \frac{1359 \times 3^{2/3}}{53900}. \quad (97)$$

The general expression (20) gives

$$T_{1,2} = \left(\frac{\pi}{2}\right)^{1/2}(2n+1)^{1/4}i^{\pm 5/6}e^{w^2/2}\frac{z^{1/6}\zeta}{\sqrt{\sin\theta}}\,H_{1/3}^{(1,2)}(\eta),$$

where $|\arg\eta| < \pi$. \hfill (98)

Substituting (98) and (82) in (77) we shall have $V_n(w)$. Expressions for $W_n(w)$ and $U_n(w)$ can be obtained in a similar way. Thus, we have finally

$$\left.\begin{aligned} U_n(w) &= 2\{J_{1/3}(\eta) + J_{-1/3}(\eta)\}S(n, w)\\ V_n(w) &= -i^{-1/3}H_{1/3}^{(2)}(\eta)S(n, w)\\ W_n(w) &= -i^{1/3}H_{1/3}^{(1)}(\eta)S(n, w) \end{aligned}\right\}, \quad (99)$$

where $|\arg n| < \pi$, $|\arg \eta| < \pi$, and

$$S(n, w) = \left(\frac{\pi}{2}\right)^{1/2}\frac{(2n+1)^{1/2}\exp\left\{\dfrac{1}{8(2n+1)^2}\right\}}{2\Gamma\left(\dfrac{n}{2}+1\right)}\frac{e^{w^2/2}z^{1/6}\zeta}{\sqrt{\sin\theta}}.$$

$$\quad (100)$$

5. Coulomb Wave Functions

Coulomb wave functions F_L and G_L are defined as the solutions of the differential equation

$$\frac{d^2y}{d\rho^2} + \left\{1 - \frac{2\eta}{\rho} - \frac{L(L+1)}{\rho^2}\right\}y = 0 \quad (101)$$

such that as $\rho \to \infty$

$$F_L(\rho, \eta) \sim \sin(\rho - \eta\log 2\rho - L\pi/2 + \sigma_L), \quad (102)$$

$$G_L(\rho, \eta) \sim \cos(\rho - \eta\log 2\rho - L\pi/2 + \sigma_L), \quad (103)$$

where $\sigma_L = \arg\Gamma(L+1+i\eta)$. Recently a number of investigations concerning their asymptotic behavior have been made, but the various formulas so far obtained do not seem to be quite satisfactory in view of their involved construction.

For simplicity, we shall here consider $F_0(\rho, \eta)$ and $G_0(\rho, \eta)$, so that (101) reduces to

$$\frac{d^2y}{d\rho^2} + \left(1 - \frac{2\eta}{\rho}\right)y = 0. \quad (104)$$

On putting

$$\rho = 2\eta(1+x), \quad 2\eta = a, \quad (105)$$

$$\Phi = F_0(\rho, \eta), \quad (106)$$

$$P = 1 - \frac{1}{1+x}, \quad (107)$$

(104) reduces to the normal form (6). If we put

$$x = \tfrac{1}{2}(\cosh t - 1) = \sinh^2\frac{t}{2}, \quad (108)$$

we easily find that

$$z = \tfrac{1}{2}(\sinh t - t), \quad (109)$$

$$P = \tanh^2\frac{t}{2}, \quad (110)$$

$$Q = -\frac{5+8x}{16x^3(1+x)}, \quad (111)$$

$$Q_1 = \frac{5+4x+8x^2}{24x^{3/2}(1+x)^{1/2}}. \quad (112)$$

From (107) it follows that

$$a_1 = 1, \quad b_1 = -1, \quad b_2 = 1, \quad b_3 = -1, \quad b_4 = 1. \quad (113)$$

Hence

$$\lambda = -\frac{4}{35}\left(\frac{3}{2}\right)^{1/3}, \quad \lambda_0 = \frac{16}{75}, \quad \lambda_1 = -\frac{3879}{13475}\left(\frac{2}{3}\right)^{1/3}. \quad (114)$$

By Stirling's formula we can write for $a > 0$

$$\sigma_0 = \arg\Gamma\left(1 + \frac{i}{2}a\right)$$

$$= \frac{\pi}{4} + \frac{a}{2}\left(\log\frac{a}{2} - 1\right) - \frac{1}{6a} + O(a^{-3}). \quad (115)$$

Hence

$$\rho - \eta\log 2\rho + \sigma_0 = az + \frac{\pi}{4} - \frac{1}{6a} + O(a^{-3}).$$

Accordingly, the asymptotic behavior (102) and (103) prescribed for the functions F_0 and G_0 can also be expressed as follows

$$G_0 \pm iF_0 \sim \exp\left[\pm i\left(az + \frac{\pi}{4} - \frac{1}{6a}\right)\right] + 0(a^{-3}) \quad (116)$$

as $z \to \infty$.

But, it follows from (107), (111), and (112) that $P \to 1$, $Q \to 0$, $Q_1 \to \frac{1}{3}$ as $z \to \infty$, and hence, on account of (9) and (19), (104) has particular solutions which, as $z \to \infty$, behave as

$$\Phi_{1,2} \sim \exp\left[\pm iaz \mp \frac{i}{6a}\right] + 0(a^{-3}). \quad (117)$$

Comparing this with (116), we immediately find that

$$G_0 \pm iF_0 = P^{-1/4} \exp\left[\pm iaz \pm \frac{\pi}{4}i \mp \frac{i}{2a}Q_1 \right.$$
$$\left. + \frac{1}{4a^2}Q\right] + 0(a^{-3}), \quad (118)$$

so that

$$F_0 = P^{-1/4}e^{Q/4a^2}\sin\left(az + \frac{\pi}{4} - \frac{1}{2a}Q_1\right) + 0(a^{-3}), \quad (119)$$

$$G_0 = P^{-1/4}e^{Q/4a^2}\cos\left(az + \frac{\pi}{4} - \frac{1}{2a}Q_1\right) + 0(a^{-3}) \quad (120)$$

for $z \neq 0$. On the other hand, for $\rho \doteq 2\eta$, that is, for $z \doteq 0$, we have in place of (118)

$$G_0 \pm iF_0 = \left(\frac{\pi a}{2}\right)^{1/2}P^{-1/4}i^{\pm 4/3}z^{1/6}\zeta H_{1/3}^{(1,2)}(\eta) + 0(a^{-3}). \quad (121)$$

Hence

$$F_0 = \left(\frac{\pi a}{2}\right)^{1/2}\frac{1}{\sqrt{3}}P^{-1/4}z^{1/6}\zeta[J_{-1/3}(\eta) + J_{1/3}(\eta)] + 0(a^{-3}), \quad (122)$$

$$G_0 = \left(\frac{\pi a}{2}\right)^{1/2}P^{-1/4}z^{1/6}\zeta[J_{-1/3}(\eta) - J_{1/3}(\eta)] + 0(a^{-3}) \quad (123)$$

for $z \doteq 0$.

Next, remembering (26) and (27), we find that

$$\left.\sqrt{3}\begin{array}{c}F_0(a, a/2) \\ G_0(a, a/2)\end{array}\right\} = \frac{(\pi a)^{1/2}}{3^{1/6}\Gamma(\frac{2}{3})}\varkappa^{-1/3}\left\{1 \mp \frac{2 \times 3^{1/3}}{35}\right.$$
$$\left.\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{4}{3})}\varkappa^{-4/3}\right\} + 0(a^{-3}), \quad (124)$$

and

$$\left.\sqrt{3}\begin{array}{c}F_0{}'(a, a/2) \\ G_0{}'(a, a/2)\end{array}\right\} = \left(\frac{\pi}{a}\right)^{1/2}\frac{\varkappa^{1/3}}{3^{5/6}\Gamma(\frac{4}{3})}\left\{\pm 1\right.$$
$$+ \frac{3^{2/3}\Gamma(\frac{4}{3})}{5\Gamma(\frac{2}{3})}\varkappa^{-2/3} \mp \frac{6}{175}\varkappa^{-2}$$
$$\left. + \frac{1596 \times 3^{2/3}}{67375}\frac{\Gamma(\frac{4}{3})}{\Gamma(\frac{2}{3})}\varkappa^{-8/3}\right\} + 0(a^{-3}). \quad (125)$$

Introducing $\varkappa^2 = a^2 + \lambda_0 = a^2 + 16/75$, we have

$$\left.\sqrt{3}\begin{array}{c}F_0(a, a/2) \\ G_0(a, a/2)\end{array}\right\} = \frac{\pi^{1/2}}{3^{1/6}\Gamma(\frac{2}{3})}a^{1/6}\left\{1 \mp \frac{2 \times 3^{1/3}}{35}\right.$$
$$\left.\frac{\Gamma(\frac{2}{3})}{\Gamma(\frac{4}{3})}a^{-4/3} - \frac{8}{225}a^{-2}\right\} + 0(a^{-3}), \quad (126)$$

$$\left.\sqrt{3}\begin{array}{c}F_0{}'(a, a/2) \\ G_0{}'(a, a/2)\end{array}\right\} = \frac{\pi^{1/2}}{3^{5/6}\Gamma(\frac{4}{3})}a^{-1/6}\left\{\pm\left(1 + \frac{2}{1575}a^{-2}\right)\right.$$
$$\left. + \frac{3^{2/3}\Gamma(\frac{4}{3})}{5\Gamma(\frac{2}{3})}\left(a^{-2/3} + \frac{1436}{17325}a^{-8/3}\right)\right\} + 0(a^{-3}). \quad (127)$$

Inserting the numerical values, we have finally

$$\left.\begin{array}{c}F_0(a, a/2) \\ G_0(a, a/2)\end{array}\right\} = \left\{\begin{array}{c}0.6292708 \\ 1.0899291\end{array}\right\}a^{1/6}[1 \mp 0.1249733\, a^{-4/3}$$
$$- 0.0355556\, a^{-2} + 0(a^{-10/3})], \quad (128)$$

$$\left.\begin{array}{c}F_0{}'(a, a/2) \\ G_0{}'(a, a/2)\end{array}\right\} = \left\{\begin{array}{c}0.4587454 \\ 0.7945704\end{array}\right\}a^{-1/6}[\pm 1 + 0.2743442\, a^{-2/3}$$
$$\pm 0.0012698\, a^{-2} + 0.0227393\, a^{-8/3} + 0(a^{-10/3})]. \quad (129)$$

Barfield and Broyles[11] have recently given the numerical table of $F_0(a, a/2)$, $G_0(a, a/2)$, and $F_0{}'(a, a/2)$ for $a = 5\ (2.5)\ 20;\ 20\ (5)\ 100$. But it is believed that our formulas (128) and (129) are more accurate and convenient for use than their numerical table.

Further, it should be mentioned that (122) and (123) give the values of $F_0(\rho, \eta)$ and $G_0(\rho, \eta)$ for any values of $\rho = 2\eta(1 + x)$ and η where $\rho \doteq 2\eta$, which are accurate to the same degree as $F_0(a, a/2)$ and $G_0(a, a/2)$. Quite recently Biedenharn, Gluckstern, Hull and Breit[12] have obtained the formulas for any arbitrary value of L, which agree with (126) and (127) for $L = 0$. But their expressions valid in the vicinity of the turning point ($\rho \doteq 2\eta$) are apparently much more complicated than our expressions (122) and (123).

[11] W. D. Barfield and A. A. Broyles, "Coulomb functions for heavy nuclear particles," *Phys. Rev.*, vol. 88, p. 892; November, 1952.

[12] L. C. Biedenharn, R. L. Gluckstern, M. H. Hull, Jr., and G. Breit, "Coulomb functions for large charges and small velocities," *Phys. Rev.*, vol. 97, pp. 542–554; January, 1955.

Approximate Method For Scattering Problems

C. E. SCHENSTED†

THE PROBLEM with which we are concerned in this paper is the determination of a function, Ψ, which satisfies certain boundary conditions and the equation $(\nabla^2 + k^2)\Psi = 0$ where $k = 2\pi/\lambda$ and λ is the wavelength. Now Ψ is a rapidly varying function, but most of the variation is due to the nearly uniform change of phase with distance. Thus it is reasonable to factor off the phase factor so as to be able to deal with a slowly varying function. It is not possible to factor out the exact phase factor (since it is unknown), but we can factor out the geometrical optics approximation to the phase. Thus we write $\Psi = \psi e^{iks}$ where $s(x, y, z)$ is the distance along the light rays. Making use of the fact that $(\nabla s)^2 = 1$ we obtain the equation

$$\frac{\partial \psi}{\partial s} + \tfrac{1}{2}\psi\nabla^2 s = \frac{i\lambda}{4\pi}\nabla^2\psi, \qquad (1)$$

for ψ, where $\partial\psi/\partial s = \nabla s\cdot\nabla\psi$ is the directional derivative of ψ along the light rays.

When the wavelength is sufficiently short we can, in the first approximation, neglect the right side of (1) and obtain an ordinary differential equation for ψ. This differential equation tells us that ψ increases when the rays are converging and decreases when the rays are diverging. It can be shown that the solution to the differential equation is just geometric optics. By substituting this approximation into the right side of (1) and solving the resulting equation we get a higher order approximation. If we repeat this procedure a number of times, we obtain the Luneberg-Kline series. The true solution approaches geometric optics in the limit of small wavelengths. However, even for very short wavelengths, we may desire a more accurate answer. For example geometric optics may sometimes give zero for the field in which case we would want more precise information.

From (1) we can see that geometric optics is only valid when ψ varies slowly so that $\lambda\nabla^2\psi$ may be neglected. But even when it is not valid to neglect $\nabla^2\psi$ it may be valid to neglect some terms of $\nabla^2\psi$. For example ψ frequently varies more slowly along the rays than perpendicular to the rays. Cases where this is true are the transition from light to shadow near a shadow boundary and the abrupt change in intensity near the edge of the finite piece of a plane wave reflected by a finite plane surface. In this case we can get a suitable approximation by neglecting the derivatives with respect to s in the $\nabla^2\psi$ term in (1). If we rewrite (1) in the form

$$\frac{\partial\psi}{\partial s} + \tfrac{1}{2}\psi\nabla^2 s - \frac{i\lambda}{4\pi}\left(\nabla^2\psi - \frac{\partial\psi}{\partial s}\nabla^2 s - \frac{\partial^2\psi}{\partial s^2}\right)$$

$$= \frac{i\lambda}{4\pi}\left(\frac{\partial\psi}{\partial s}\nabla^2 s + \frac{\partial^2\psi}{\partial s^2}\right) \qquad (2)$$

then this approximation consists of neglecting the right side of (2). This approximate equation is very nearly

the same as the one used by Fock in his determination of the fields near a shadow curve.

As before we can use the first approximation in the right side of (2) and so obtain a second order approximation, etc.

Eq. (2) without the right side may be given a physical interpretation which indicates the way in which the field propagates. If we move along a ray then, as before, the second term on the left shows us that ψ increases or decreases according as the rays are converging or diverging. The third term on the left shows us that, in addition, there is a diffusion (with imaginary diffusion constant) in a direction perpendicular to the rays. Thus near a shadow boundary there is a diffusion of ψ from the illuminated region into the shadow region. Since the diffusion constant is proportional to λ, this diffusion is most pronounced for long wavelengths. If we include the right side of (2) we see that there is also a diffusion along the rays.

The approximation obtained by neglecting the right side of (2), though more accurate, is not nearly so simple to compute as the approximation obtained by neglecting the right side of (1) as we still have a partial differential equation. But this equation has various advantages over the original equation, $(\nabla^2 + k^2)\Psi = 0$. One is the physical interpretation given above. Another is that the finite difference approximation to (2) (neglecting the right side) can be solved for $\psi(s + \nabla s)$ when $\psi(s)$ is given (since the equation is parabolic) while the finite difference approximation to $(\nabla^2 + k^2)\Psi = 0$ is a system of simultaneous equations (since the equation is elliptic).

In the following we will give two examples of the use of this approximation method. We will take the boundary condition to be $\psi = 0$. We will assume that there is an incident field for which $\psi_i = -1$ (this corresponds to a plane wave) and will compute only the scattered field. The scattered field satisfies the boundary condition $\psi = 1$.

The first problem we will consider is the scattering by the half-plane $y = 0$, $x < 0$ (Fig. 1), with the direction of incidence being the negative x direction. If we introduce the polar co-ordinate system ρ, ϕ through

$$x = \rho\cos\phi \qquad (3)$$
$$y = \rho\sin\phi,$$

then $s = \rho$ and the approximate differential equation is

$$\frac{\partial\psi}{\partial\rho} + \frac{\psi}{2\rho} = \frac{i\lambda}{4\pi\rho^2}\frac{\partial^2\psi}{\partial\phi^2}. \qquad (4)$$

The boundary condition is

$$\psi(\rho, \pm\pi) = 1. \qquad (5)$$

If we use separation of variables with $\psi(\rho, \phi) = R(\rho)\Phi(\phi)$ we get

$$\frac{4\pi\rho^2}{i\lambda}\left(\frac{R'}{R} + \frac{1}{2\rho}\right) = \frac{\Phi''}{\Phi} = \mu^2, \qquad (6)$$

† University of Michigan, Ann Arbor, Mich.

Fig. 1

Fig. 2

(a)　　　　　　　(b)　　　　　　　(c)
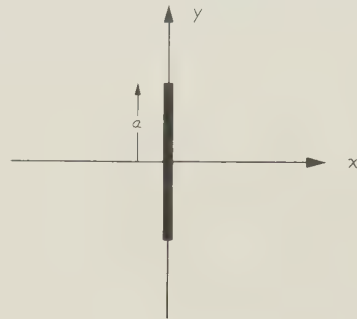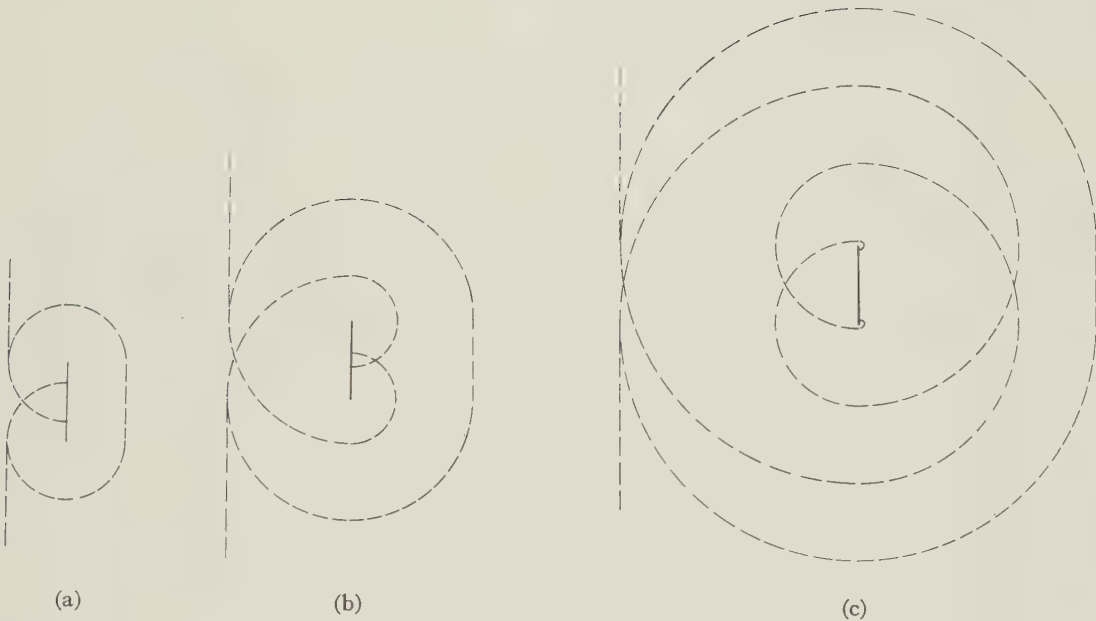
Fig. 3

so that

$$\Phi = A \cosh \mu\phi + B \sinh \mu\phi \qquad (7)$$

$$R = \frac{1}{\sqrt{\rho}} e^{-\frac{i\lambda}{4\pi\rho} \mu^2}.$$

Since $\Phi(\pi) = \Phi(-\pi)$ we have $B = 0$. Thus we look for a solution in the form

$$\psi = \int_{-\infty}^{\infty} A(\mu) \frac{1}{\sqrt{\rho}} e^{-\frac{i\lambda}{4\pi\rho} \mu^2} \cosh \mu\phi \, d\mu, \qquad (8)$$

where $A(\mu)$ is determined from (5). Now

$$\int_{-\infty}^{\infty} e^{-\frac{i\lambda}{4\pi\rho} \mu^2} d\mu = 2\pi \sqrt{\frac{\rho}{\lambda}} \, e^{-\frac{\pi i}{4}}, \qquad (9)$$

so that

$$A(\mu) = \frac{\sqrt{\lambda}}{2\pi} e^{\frac{\pi i}{4}} \frac{1}{\cosh \mu \pi}.$$

Thus

$$\psi = \frac{1}{2\pi} \sqrt{\frac{\lambda}{\rho}} e^{\frac{\pi i}{4}} \int_{-\infty}^{\infty} e^{-\frac{i\lambda}{4\pi\rho} \mu^2} \frac{\cosh \mu\phi}{\cosh \mu\pi} \, d\mu. \qquad (10)$$

For very large ρ (large enough that

$$- (\pi - \phi) \sqrt{\frac{4\pi\rho\epsilon}{\lambda}}$$

$$e \qquad\qquad << 1,$$

where ϵ is a fixed number which is determined by the desired accuracy) we have

$$\psi \approx \frac{1}{2\pi} \sqrt{\frac{\lambda}{\rho}} e^{\frac{\pi i}{4}} \int_{-\infty}^{\infty} \frac{\cosh \mu\phi}{\cosh \mu\pi} \, d\mu = \frac{e^{\frac{\pi i}{4}}}{\sqrt{2\pi k\rho} \cos \frac{\phi}{2}}. \qquad (11)$$

This is exactly the same result as that given by the exact (Sommerfeld) half-plane solution for large ρ. For the same problem geometric optics gives $\psi = 0$.

The second problem which we will consider is the scattering by a circular disk. In this case we will make some additional approximations. Again the incident field will be taken to be going in the negative x direction (Fig. 2). The radius of the disk will be taken to be a. In this problem the wave fronts will have the shape shown in Fig. 3. Figs. 3(a), 3(b), and 3(c) are for $s = 1.45a$, $s = 3.1a$, and $s = 6.15a$ respectively. We interest ourselves in the value of ψ only for $r = \sqrt{y^2 + z^2}$ $= 0$.

On the flat reflected part of the wave front $\psi = 1$ for $s = 0$. On the curved diffracted part of the wave front geometric optics would give $\psi = 0$ so that ψ differs from zero only as a result of the diffusion. Furthermore the ψ that diffuses into the diffraction region is reduced due to the divergence of the rays. Thus the reflected part of the wave front acts as a source of ψ and the diffracted parts act as a sink. When the wavelength is short enough that ψ does not diffuse appreciably from the forward direction all the way to the backward direction we can approximate the effect of the sink, as far as the value of ψ at $r = 0$ is concerned, by imposing the boundary condition $\psi = 0$ at $r = a$.

On the reflected rays, when we neglect the right side, (2) becomes

$$\frac{\partial \psi}{\partial x} = \frac{i\lambda}{4\pi r} \frac{\partial}{\partial r}\left(r \frac{\partial \psi}{\partial r}\right). \tag{12}$$

We have $\psi(0, r) = 1$ and $\psi(x, 0) < \infty$. Also we are imposing the condition $\psi(x, a) = 0$. If we let $L(p, r)$ be the Laplace transform of ψ, then the transform of (12) is

$$\frac{\partial^2 L}{\partial r^2} + \frac{1}{r} \frac{\partial L}{\partial r} + \frac{4\pi i p}{\lambda} L = \frac{4\pi i p}{\lambda}. \tag{13}$$

The solution of (13) satisfying the boundary conditions is

$$L = 1 - \frac{J_o\left(\sqrt{\frac{4\pi p}{\lambda}} \, e^{\frac{\pi i}{4}} \, r\right)}{J_o\left(\sqrt{\frac{4\pi p}{\lambda}} \, e^{\frac{\pi i}{4}} \, a\right)}. \tag{14}$$

Due to the expansion

$$1 - \frac{J_o\left(\sqrt{\frac{4\pi p}{\lambda}} \, e^{\frac{\pi i}{4}} \, r\right)}{J_o\left(\sqrt{\frac{4\pi p}{\lambda}} \, e^{\frac{\pi i}{4}} \, a\right)} = 1 - \frac{1 - \frac{\pi i p}{\lambda} r^2 + \cdots}{1 - \frac{\pi i p}{\lambda} a^2 + \cdots}$$

$$= \frac{\pi i p}{\lambda} (r^2 - a^2) + \cdots, \tag{15}$$

we find that for large x

$$\psi(x, 0) \approx - \frac{\pi i a^2}{\lambda x}. \tag{16}$$

This is identical with the formula which would be obtained by applying physical optics to this problem. It would seem that if we made an assumption which is not as crude as assuming $\psi(x, a) = 0$ that we would have obtained a better result than (16).

Electromagnetic Research at the Institute of Mathematical Sciences of New York University[*]

MORRIS KLINE[†]

Summary—This paper presents current electromagnetic research efforts and research completed during the past few years at New York University. In the research itself emphasis has been placed on basic problems involving appreciable mathematical complexity and mathematical methodology. However, this account describes the results from the standpoint of their contribution to microwave problems, ionospheric and tropospheric propagation, diffraction, the inverse propagation and synthesis problem, antenna and waveguide theory, and other physical problems.

INTRODUCTION

I WELCOME the opportunity afforded by the sponsors of this Symposium to describe the research in electromagnetic theory performed at the Institute of Mathematical Sciences of New York University. It permits us to "take stock," a wise measure for the scientist as well as the business man.

It has been our conviction for a number of years that electromagnetic problems should be one of the special concerns of mathematicians. Maxwell very kindly bequeathed to posterity some amazingly exact basic laws but, unfortunately, did not show how to solve the boundary value problems which arise in the applications of his laws. However over the past 75 years a number of mathematicians and physicists, one thinks immediately of Rayleigh, Sommerfeld, Weyl, Watson, van der Pol, and Bremmer, solved some basic and practically important problems. A survey of their work shows, on the one hand, that mathematical proficiency and diligence can produce solutions and, on the other, that the problems are sufficiently difficult, and indeed becoming more and more so, to require the services of skilled mathematicians. Hence, with full and conscious recognition of the reservations which must be made as to our skill, we at New York University have been tackling electromagnetic problems in which mathematical obstacles have to be surmounted.

Though certain classes of problems such as diffraction theory, tropospheric and ionospheric propagation, and methodological studies have been pursued rather consistently, we have also subscribed to the principle that freedom in research is essential. Hence we have often found ourselves with results that willed to be solved rather than with results that we necessarily willed to solve. As a consequence I am obliged to offer you a smorgasbord rather than a planned meal. I trust that the offering will be none the less appetizing.

As I have indicated, we have approached a variety of problems from the standpoint of the mathematical difficulties involved but, since the interests of this audience no doubt center on the electromagnetic content of our work, I shall endeavor to organize our results in accordance with their physical significance. Appended to this paper are the titles of several series of reports we have issued during the past nine years. References in the paper are to these lists. Space and time will not permit full recognition of a number of related contributions from research men outside of New York University to many of whom we are deeply indebted. However our individual papers generally give the appropriate connections and references.

FIELD SOLUTIONS IN THE QUASI-OPTICAL RANGE

The rapid development of ultra-high frequency and microwave techniques has given rise to hundreds of new electromagnetic problems. These frequencies are, of course, close to the optical range and, therefore, electromagnetic theorists have utilized geometrical optics methods, which were extensively explored during the two hundred years before the birth of electromagnetic theory. The solutions thereby obtained are necessarily approximate and often unsatisfactory. Hence, from the very outset of our work we have sought a systematic attack on electromagnetic problems which would take advantage of the existing optical methods and yet improve on them in the direction of wave theory; that is, we have sought to develop the transition from optics to electromagnetics. The link between these two domains, which has proved to be both theoretically and practically significant, is the notion, due to Luneberg, of asymptotic solution of electromagnetic problems (EM-14, 15). In this approach one seeks not exact solutions of electromagnetic problems but asymptotic series solutions in which the basic variable in the series is the wavelength λ or the reciprocal of the circular frequency ω of a monochromatic source.

The first major step in attempting an asymptotic series solution of an electromagnetic problem is to determine the form of the asymptotic series which is valid for the particular type of problem, e.g., reflection, diffraction, or refraction. The second step is to learn how to calculate these series. To accomplish the first objective we have utilized the concept of the pulse solution of an electromagnetic time harmonic problem. The pulse field corresponding to an electromagnetic field due to a finite aggregate of charges or charges at infinity having a harmonic time behavior $e^{-i\omega t}$ is the field due to the very same charges but having a Heaviside unit function time behavior, that is, O for $t < O$ and 1 for $t > O$. By means of the relationship between the pulse solution and the steady-state time harmonic solution we have

† New York University, New York, N. Y.

derived the proper form of the asymptotic series in those spatial regions where reflection only (in the geometrical optics sense) takes place and in certain regions of diffraction, notably, the regions into which fields diffracted by an aperture or an optical lens system penetrate. In the cases of directly transmitted and reflected fields the asymptotic series which give the spatial behavior of E and H (the time behavior is $e^{-i\omega t}$) consist essentially of terms in the zeroth and positive integral powers of $1/\omega$ (each multiplied by a phase factor) and in the case of aperture diffraction fields the series contain positive fractional powers of $1/\omega$.

The asymptotic form of the fields on and near foci and caustics of certain reflecting surfaces, e.g., cylindrical reflectors, has been determined by use of the above results and Green's theorem (EM-55). The results give the amplitude and phase of the asymptotic solution at any point on a caustic in terms of the curvatures of the incident wave front and the reflecting surface and of the angle of incidence. On the caustic the leading term in the asymptotic series is proportional to $k^{1/2-1/n}$ where $k = 2\pi/\lambda$ and n is an integer greater than 2 whose value depends upon the equation of the caustic. At the focus of a parabolic cylinder the leading term is proportional to $k^{1/2}$.

The asymptotic form of fields diffracted by obstacles in homogeneous media such as cylinders or spheres in free space has posed a more difficult problem. At the present time our results limit the forms which obtain in such fields (EM-67). The chief characteristic of these forms is an exponential decay factor and fractional exponents in the series for $1/\omega$. The exponential factor (apart from a phase factor) is e^{-kax} where $0 < a \le \frac{1}{2}$ and $\chi(x, y, z) = $ const are certain determinable equi-amplitude surfaces. The fractional powers of $1/\omega$ which appear in the series depend upon the value of a which is needed for a particular problem and are determinable in terms of a. If, for example, $a = 1/2$ then the powers of $1/\omega$ are of the form $(1/\omega)^{n/2}$. If $a = 1/3$, the powers of $1/\omega$ are of the form $(1/\omega)^{n/3}$. For other values of a, the powers are not simply stated.

From the theory of these forms of the asymptotic series (EM-24, 48, 67) one finds that the coefficients satisfy certain linear first order ordinary differential equations if one follows the behavior of these coefficients along lines or curves orthogonal to surfaces of equal phase. These surfaces in turn satisfy the eiconal equation of optics, namely, $(\nabla\phi)^2 = \epsilon\mu$, and are the wave fronts. The family of orthogonal curves is then the family of rays of geometrical optics. Since the theory of wave fronts and rays has been developed only for the phenomena of reflection and refraction (see also EM-13, 20) it is necessary to extend this theory to diffracted fields. The chief features of this extension are as follows.

If a plane wave is incident upon a cylinder of bounded cross section the rays of the incident field are just the normals of the plane-wave front. Some of these proceed directly past the obstacle. Others are reflected according
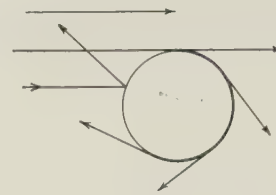
Fig. 1

to the usual laws of geometrical optics. However, the two rays which strike the surface tangentially wind around the surface and shed rays tangential to the surface as they travel around (Fig. 1). Each of the two tangential rays travels around the cylinder an infinite number of times.
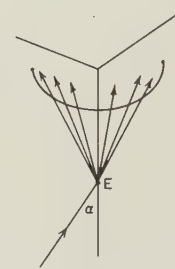
Fig. 2

In the case of rays striking an edge, for example of a perfectly conducting wedge, at an angle a, a partial cone of rays is diffracted with the edge as axis and semi-vertex angle a (Fig. 2). If the wedge is a dielectric then rays enter the dielectric at E and fill out another partial cone interior to the wedge with vertex at E and semi-vertex angle a' determined by Fresnel's law of refraction.

Fig. 3

A ray striking a corner (C in Fig. 3) generates diffracted rays which spread out in every direction from the corner. A ray striking the edge of an aperture normal to the surface containing the aperture generates diffracted rays in every direction in a plane containing the incident ray and normal to the aperture curve at the point of incidence (Fig. 4).

In all of the above cases if one takes a point P on the incident ray and a point Q on a diffracted ray then the optical path from P to Q renders Fermat's integral stationary under the appropriate boundary condition that the path must include a point or curve on the diffracting surface.

As noted above, the individual coefficients of the asymptotic series each satisfies a first order linear
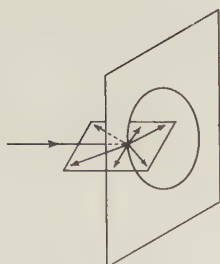
Fig. 4

ordinary differential equation in which the independent variable is arc length or some other suitable parameter along the ray. Moreover, the entire set of ordinary differential equations forms a recursive system. The nonhomogeneous term of any one differential equation is known in terms of the solutions of the preceding equations. To obtain the unique solutions of these differential equations initial values are needed. These can be obtained for many classes of problems but there are still difficulties in, so to speak, starting the asymptotic series along rays in the diffracted region.

One other way of obtaining the asymptotic series has been found to be useful in problems of diffraction by an aperture or by an optical lens system. The theoretical work on the relation between the pulse solution and the time harmonic solution of an electromagnetic problem has led to a rather broad theorem which states in effect that the singularities of the pulse solution (the discontinuities of the pulse solution or the discontinuities of the first or higher order time derivatives) determine completely the asymptotic series representation of the steady-state time harmonic field (EM-15 and current work). If the behavior of the pulse solution in the neighborhood of a singularity is known then one can write down at once the contribution to the asymptotic series due to that singularity of the pulse solution. This theorem has been applied to obtain the asymptotic form of well known double integrals which give the diffracted fields of rather general, idealized optical lens systems. The pulse solution of such a problem can be written down at once if the geometrical optics behavior of the lens system is known. Analysis of the singularities of the pulse solution then yields the asymptotic form of the diffracted field. The method involved here is essentially stationary phase extended to double integrals of functions of two real variables.

In view of our interest in pulse solutions in behalf of the theory of asymptotic solution of electromagnetic problems we have solved a number of pulse problems. The specific results will be discussed under the subjects of propagation and diffraction where these results have independent significance and application.

A number of practical applications of the above theory have been made. Plane, circularly cylindrical, and spherical waves have been treated by the above methods with the idea of obtaining general asymptotic formulas for waves having these geometrical characteristics.

Typical results include, for example, the determination of the field due to a line source at the focus of a perfectly conducting parabolic cylinder or a point source at the focus of a paraboloid of revolution on which the field or its normal derivative vanishes. Also included are the fields produced by the same surfaces when plane waves are incident from the other (convex) side along the respective axes of symmetry. These results have been checked against the asymptotic evaluation of the exact solutions of those diffraction problems for which exact solutions are known.

Propagation in the Ionosphere

Though the effect of the ionosphere on electromagnetic wave propagation has been studied for fifty years, definitive results are lacking. The reason is that the ionosphere in the presence of the Earth's magnetic field presents a nonhomogeneous anisotropic medium with several parameters such as collision frequency, the gyromagnetic frequency of the electrons, and density of the electrons, each varying as a function of position and time, about which quantitative knowledge is fragmentary. Even if the parameters and consequently the dielectric tensor were known precisely the mathematical problem of predicting the behavior of a plane wave or dipole field which enters the ionosphere from below is still beyond the full powers of present-day mathematics. Awareness of this fact has led us to undertake fundamental theoretical work bearing on this problem.

On the assumption that the ionosphere is horizontally stratified so that the dielectric tensor varies in one direction only, Maxwell's equations applied to this anisotropic medium can be reduced to a system of four first-order linear ordinary differential equations in which four of the six components of E and H are the dependent variables. Hence we have undertaken to study the theory of such systems of differential equations. This theory had not been advanced very far beyond the derivation of existence and uniqueness theorems. Even the analogs of some of the many artifices which allow one to simplify or to standardize second order ordinary differential equations for one unknown function have hardly been considered. From the algebraic point of view, the most obvious line of approach seems to be to attempt to generalize the exponential form of the solution of a linear differential equation for a single unknown function. (Problems related to this one also play a role in the quantum theory of fields.) The purely algebraic aspects of the investigations to be described apply also to linear operators in Hilbert space.

We write (BR-3) a system of n homogeneous linear differential equations for n unknown functions in the form

$$\frac{dY(t)}{dt} = A(t)Y(t), \tag{1}$$

where $A(t)$ is an n by n matrix the elements of which are functions of the real variable t and where $Y(t)$ is an unknown n by n matrix the column vectors of which

form a set of linearly independent solutions of the system of linear differential equations

$$\frac{dy_v}{dt} = \sum_{\mu=1}^{n} a_{v,\mu}(t)y_{\mu}, \qquad v = 1, 2, \cdots, n. \qquad (2)$$

In (2) the $a_{v,\mu}$ are the elements of A and $y_1, \cdots, y_n$ are scalar functions of t.

We now try to find an n by n matrix $\Omega(t)$ such that

$$Y = \exp \Omega(t) = I + \Omega + \frac{\Omega^2}{2!} + \cdots \qquad (3)$$

is a solution of (1) satisfying $Y(O) = I$, where I denotes the identity matrix. For sufficiently small values of t, $\Omega(t)$ is given by an infinite series the first terms of which can be determined explicitly, viz:

$$\Omega = \int_o^t A(\tau) \, d\tau + \frac{1}{2} \int_o^t \left[A(\tau), \int_o^\tau A(\sigma) \, d\sigma \right] d\tau$$

$$+ \frac{1}{4} \int_o^t \left[A(\tau), \int_o^\tau \left[A(\sigma), \int_o^\sigma A(\rho) \, d\rho \right] d\sigma \right] d\tau$$

$$+ \frac{1}{12} \int_o^t \left[\left[A(\tau), \int_o^\tau A(\sigma) \, d\sigma \right], \int_o^\tau A(\sigma) \, d\sigma \right] d\tau + \cdots . \qquad (4)$$

In (4) the square brackets [ ] denote Lie multiplication; that is, for any two $n \times n$ matrices L and M, $[L, M] = LM - ML$. The right-hand side of (4) has the following property: If iA is an Hermitian matrix then iX is again Hermitian for every term X on the right-hand side of (4). Formula (4) is the continuous analog of the Baker-Hausdorff formula (i.e., of the addition theorem of the exponential function in the case of noncommutative variables).

The range of validity of (4) is a finite one in most cases. Sufficient conditions for the validity of this formula for all values of t have been derived. A case in which (4) is certainly valid for all values of t arises if the series on the right-hand side terminates. In this case, the formula provides a solution of (1) in finite terms (quadratures, exponential functions and algebraic operations), because $\exp \Omega$ can be written as follows:

$$\exp \Omega = \sum_{v=1}^{n} \exp \left(\lambda_v \frac{P_v(\Omega)}{P'(\lambda_v)} \right), \qquad (5)$$

where the λ_v are the eigenvalues of Ω and

$$P_v(\lambda) = \frac{P(\lambda)}{\lambda - \lambda_v} \qquad (6)$$

is a polynomial in λ which is obtained from the characteristic equation $P(\lambda) = 0$ for Ω by (6).

We have investigated the case where

$$\Omega(t) = \int_o^t A(\tau) \, d\tau \qquad (7)$$

or, equivalently,

$$\left[A(t), \int_o^t A(\tau) \, d\tau \right] = 0 \qquad (8)$$

and have found $(BR\text{-}10)$ that (8) may hold even though it is impossible to find a constant matrix T such that $T^{-1}A(t)T$ is triangular for all t. This shows that (8) is a nontrivial test for the existence of solutions in finite terms. More general criteria for solution of the system in finite terms have also been derived.

The work on this theory is continuing with the general objective of obtaining additional criteria for the solution of the system in finite terms. It is hoped that models of the ionosphere can be obtained which satisfy one or more of these criteria and thereby permit a fuller and deeper investigation of propagation in this medium.

An alternative, and perhaps less ambitious, approach to the problem of solving a system of n homogeneous linear ordinary differential equations has been investigated. In this approach a full infinite series expansion of the solution is involved and retained as such (EM-33). The system consists of the same n linear ordinary differential equations treated above but is restricted to self-commuting matrices, i.e., $A(x')A(x'') = A(x'')A(x')$, a condition satisfied by the matrices involved in ionospheric propagation. We shall write the system as[1]

$$\frac{dU(z)}{dz} = A(z)U(z), \qquad U(z_o) = U_o, \qquad (9)$$

in which $U(z)$ is an n-rowed column vector and $A(z)$ is an $n \times n$ square matrix whose elements are functions of real or complex z. (When the elements of A are constants the general solution is, of course, the usual linear combination of n exponential functions.)

This research starts with the Peano-Baker matrizant representation of the solution of (9). The matrizant of a matrix $A(z)$ is defined as

$$\Omega_{z_o}^z \left\{ A(x) \right\} = I + \int_{z_o}^z A(x_1) \, dx_1$$

$$+ \int_{z_o}^z A(x_1) \int_{z_o}^{x_1} A(x_2) \, dx_2 \, dx_1 + \cdots .$$

It can be shown that

$$\Omega_{z_o}^z \left\{ A(x) \right\} = \exp \int_{z_o}^z A(x) \, dx.$$

The Peano-Baker matrizant solution of the system (9) is given by

$$U(z) = \Omega_{z_o}^z \left\{ A(x) \right\} U_o .$$

However the solution converges slowly and hence is transformed. When the eigenvalues of $A(z)$ are distinct then a matrix $P(z)$ can be found such that

$$\Omega_{z_o}^z \left\{ A(x) \right\} = P(z) \, \Omega_{z_o}^z \left\{ P^{-1}(x)A(x)P(x) \right.$$

$$\left. - P^{-1}(x) \frac{dP(x)}{dx} \right\} P^{-1}(z_o) \qquad (10)$$

---

[1] We shall use z here for the independent variable for the sake of the application which follows.

in which the product $P^{-1}AP$ is a diagonal matrix. The solution is then given by

$$U(z) = P(z) \left(\exp \int_{z_0}^{z} L(x)\,dx \right) \begin{array}{c} z \\ \Omega \\ z_0 \end{array} \left\{ M(z_0, x) \right\}$$
$$\cdot P^{-1}(z_0) U_0, \quad (11)$$

where $L(x)$ is a certain diagonal matrix and $M(z_0, z)$ is a matrix which results from the separation indicated in (10). This form of the solution converges more rapidly. By further transformation it is possible to write any component $u_i(z)$ of $U(z)$ as

$$u_i(z) = \sum_{i=1}^{n} \left[\frac{p_{ij}(z) e^{\int_{z_0}^{z} g_{jj}(x)\,dx}}{\det P(z)} \right] e^{\int_{z_0}^{z} \lambda_j(x)\,dx}$$
$$\cdot C_j(z_0, z) \det P(z_0). \quad (12)$$

The main feature of this form is that each component of the solution is a sum of exponential functions with variable exponents and variable coefficients. C_j is one component of a column matrix and is itself a sum of matrizants and hence a sum of a series of multiple integrals.

If the matrix $A(z)$ contains k, where $k = 2\pi/\lambda$, in such a form that $A(z, k) = kA(z)$ then integration by parts applied to the above solutions yields an asymptotic series in positive integral powers of $1/k$. Hence both convergent and asymptotic series representations of $u_i(z)$ are obtained in this work.

When the above result is specialized to a system of two first-order ordinary differential equations, equivalent to one second order ordinary differential equation, the first term in the representation is equivalent to the usual WKB approximation. Thus the above theory is a generalization of the WKB method to systems of ordinary differential equations. The representation fails when two or more latent roots or eigenvalues of $A(z)$ are equal, which points correspond to turning points in the usual theory of second-order equations. Connection formulas are derived to connect solutions above and below turning points.

The preceding theory has been applied directly to ionospheric propagation (EM-56). The ionosphere is assumed to have the usual distribution $N(z)$ of electrons; and a static earth's magnetic field, which may vary with z, is assumed to be present. The dielectric property of the ionosphere is then described by a tensor $K(z)$ and Maxwell's equations reduce to a system of four second-order partial differential equations in the transverse components E_1, E_2, H_1, and H_2 of the vectors E and H. By assuming that E and H have the forms

$$E = V(z) \exp\left[i\frac{\omega}{c} (px + qy) \right],$$

$$H = I(z) \exp\left[i\frac{\omega}{c} (px + qy) \right],$$

the system of partial differential equations reduces to a system of four first-order ordinary differential equations in the components V_1, V_2, I_1 and I_2 of V and I. The system takes the form

$$\frac{dW(z)}{dz} = i\frac{\omega}{c} A(z) W(z),$$

which is precisely the form treated above.

As in the general theory, the introduction of a matrix $P(z)$ converts the system into

$$\frac{dU}{dz} = \left[i\frac{\omega}{c} P^{-1}(z) A(z) P(z) = P^{-1}(z) \frac{dP}{dz} \right] U(z).$$

If the matrix in the brackets could be diagonalized by the proper choice of $P(z)$ the solution $U(z)$ would then consist of the four $u_i(z)$, where $u_i(z)$ would be the exponential integral of the i-th diagonal element of the diagonalized matrix. One then has the usual characteristic waves considered in ionospheric theory. However, for a general matrix $A(z)$ whose elements are functions of z this diagonalizing process cannot be carried out. However, one can choose $P(z)$ so that $P^{-1}AP$ is diagonal if the eigenvalues of A are distinct, that is, if

$$| \lambda I - A(z) | = 0$$

has no repeated roots. This equation is the well-known fourth degree equation of magneto-ionic theory. The final form of the solution is then shown to be

$$u_j(z) = \exp\left[\left[\int_{z_0}^{z} r_{jj}(x)\,dx \right] \cdot \exp\left[\frac{i\omega}{c} \int_{z_0}^{z} \lambda_j(x)\,dx \right] \right.$$
$$\cdot C_j(z_0, z), \quad j = 1, 2, 3, 4, \quad (13)$$

where

$$C_i(z_0, z) = \sum_{j=1}^{4} \Omega_{ij}(z) u_j(z_0),$$

and where Ω_{ij} is an element of a rather complex matrizant.

What is significant about this solution is that each wave consists of three factors; the first is the usual WKB solution for the characteristic waves; the second factor contains the wave behavior; and the third factor contains the continuous interaction or coupling among four waves. Two of these four waves may be identified in special cases as the up-going and down-going ordinary waves and the other two, as the extraordinary waves. In the general case these four waves are coupled with each other continuously. In special cases there are various types of uncoupling. It follows from what was said earlier that any one of the components of the true E and H in the ionosphere is a linear combination with variable coefficients of the four characteristic waves. The full solution of the problem of a plane wave incident from below and satisfying the radiation condition above the layer is carried out in this paper.

If $P^{-1}AP$ cannot be diagonalized then A has repeated characteristic roots or eigenvalues and special reflection levels are introduced, which are analyzed. In principle the solutions may be continued past reflection levels

because they are exact rather than asymptotic solutions.

The chief significance of the above paper at the present time is to indicate the true complexity of electromagnetic wave propagation in anisotropic inhomogeneous media as opposed to the rather specialized cases treated in the literature.

A number of special studies of the anisotropic problem under restrictive assumptions have been made. In the case of low frequency vertical (normal to the earth) propagation losses in the ionosphere cannot be neglected. For vertical magnetic field, oblique incidence, and electron density exponentially increasing with height, Heading and Whipple were able to solve the system of four equations. An attempt was made (EM-51) to solve the corresponding set of equations for vertical incidence and oblique magnetic field, but the exact solution was not obtained. However, both in our case and in Heading and Whipple's case one has to assume, in order to analyze the final results, that the collision frequency is either large or small. A comparison of theoretical and experimental results shows that neither assumption is justified in the 15–150 kc range.

It is, of course, possible to make more progress in the theory of ionospheric propagation if one neglects the earth's magnetic field so that the ionosphere may be treated as an isotropic medium. This magnetic field can be neglected in the case of low frequency horizontal propagation, that is, along the earth's surface. In this case the problem becomes an eigenvalue problem. One objective of current research has been the time of arrival of the waves.

We consider first the case of horizontal polarization and zero collision frequency. The electron density is assumed to be identically zero near the surface of the earth and to increase monotonically with height. (Past investigators of this problem generally assumed constant density and a sharp transition from the lower atmosphere to the ionosphere.) Also, the frequency is assumed to be low enough to make a ray treatment of the problem (consecutive reflections between the earth and the ionosphere) inadequate. It became clear from the investigation that, except near the antipode of the transmitter, the sphericity of the earth has a negligible effect; hence the earth was assumed to be flat.

The Hertz vector u satisfies the wave equation

$$u_{rr} + u_{zz} + [k^2 - \alpha N(z)] u = 0,$$

where $N(z)$ is the electron density distribution in the vertical direction z, r is the horizontal distance from the transmitter, and α is a constant. To find the modes of propagation we solve this equation by separation of variables. Assuming that the eigenvalues k_n^2 of the ordinary differential equation involving z are known, the separation constant Γ^2 is a simple function of the free-space wave number k, i.e., $\Gamma_n^2 = k^2 - k_n^2$. We see that, just as in the case of a parallel-plane waveguide, each mode has a cutoff wave number k_n, and that at any frequency there will be a finite number of propagated

modes. The time required for the signal to travel in the n-th mode from the transmitter to a point at distance r is obtained by differentiating the phase of this expression with respect to angular frequency, thus obtaining

$$\tau = r/c \sqrt{1 - (k_n/k)^2}.$$

Locally the field at the ground can be seen to consist of a plane wave incident at the angle $\sin^{-1}(k_n/k)$ with the ground and the corresponding plane wave reflected from the ground. Thus the cosine of the angle of arrival is equal to the ratio of free-space propagation time r/c in each mode to time of travel.

In the case of horizontal polarization and exponential variation of electron density and collision frequency the eigenvalues satisfy a transcendental equation involving Bessel functions. For sufficiently low frequency the propagation constants of the first few modes can be approximated by explicit functions of frequency and the parameters of the problem. Hence the attenuation coefficient and group velocity are readily calculated for these modes.

In another of the investigations based on an inhomogeneous isotropic horizontally stratified ionosphere and wave theory we have considered vertical propagation and sought the reflection time of a pulse having carrier frequency ω_o (EM-27). We solved the monochromatic steady-state problem and calculated the complex reflection coefficient. The time of travel of the pulse is then approximately equal to the derivative of the phase of the reflection coefficient with respect to the frequency if (1) the magnitude of the reflection coefficient varies slowly with the frequency and (2) internal reflections within the ionospheric layer are negligible. (The latter assumption has not been generally recognized.) If both assumptions are satisfied and if the electron density varies sufficiently slowly so that the WKB approximation can be used to calculate the reflection coefficient we obtain an expression for the time delay (reflection time) identical with that calculated by others on the basis of group velocity considerations.

The bulk of all investigations in ionospheric wave propagation has utilized geometrical optics methods. On this basis we showed in an early work (EM-17) how to simplify the calculation of the virtual height of an ionospheric layer as a function of frequency for a vertically incident pulse by utilizing the Gauss-Christoffel integration method. More recently in work about to be published we have improved on the theory for calculating the ion distribution as a function of height by utilizing virtual height data. The anisotropy of the ionosphere is retained. The virtual height of reflection Z_v, of a pulse as a function of R (ratio of operating frequency to critical frequency) is related to the ionization distribution function $N(Z)$ through the integral equation

$$Z_v(R) = \int_0^{N_0(R)} Z'(N) \frac{\partial \mu R}{\partial R} \, dN,$$

where $Z(N)$ is the inverse of $N(Z)$, $\mu(Z)$, the index of refraction of the layer, is a function of N and of R, Z_o, the actual reflection height, is given by $\mu(Z_o) = 0$, and $N_o = N(Z_o)$. Because of the anisotropy of the ionized layer induced by the earth's magnetic field, there are two indices of refraction corresponding to the ordinary and extraordinary rays and correspondingly two virtual height curves.

Two procedures have been employed to convert the experimental virtual height data into ionization distributions. In the first, assumed distributions of the ion density as a function of height, e.g., a parabolic variation, and a known exact or approximate form of μ as a function of N have been substituted; the above integral is evaluated exactly; and certain parameters in the distribution are determined. The second procedure is to assume a practical form for the function $\mu(N)$ and to solve the integral equation for $Z(N)$. A number of cases, e.g., propagation of the ordinary ray and extraordinary ray, longitudinal propagation (along the direction of the earth's magnetic field) and transverse propagation (perpendicular to the direction of the earth's magnetic field) have been treated.[2]

TROPOSPHERIC WAVE PROPAGATION

Attention was called during World War II to marked refractive and duct effects of the atmosphere on the propagation of uhf waves. It was also gradually realized that ducts play a role even at lower frequencies. The normal mode theory used before World War II by van der Pol and Bremmer for homogeneous atmospheres was extended during the war to uhf waves in horizontally stratified media and reasonably useful predictions were made possible. However, the extended theory utilized a flat earth and a modified index of refraction which was intended to compensate for the earth-flattening assumption. This theory proved useful in the diffraction zone but calls for the summation of a large number of modes in the lit region and particularly near the horizon. The eigenvalues and eigenfunctions themselves are rather hard to calculate. The continuous spectrum (branch-cut integral) was ignored. In addition, it has been found that the first mode (eigenfunction) alone, which was used to give the field in the diffraction zone, does not suffice to account for the experimental fields obtained during the past few years in the far regions of diffraction zone. It therefore seemed desirable that the necessarily hasty and approximate work done under the pressure of the war emergency be reexamined and improved.

We have made several studies of the kinds of modes and eigenvalues that can occur in cylindrical and spherical structures. In general these studies determine the complete sets of vector orthogonal functions (eigenfunctions or modes) and their use in the transformation of vector electromagnetic problems into formally

solvable ordinary differential equations (EM-28, 29, 69, and current). The mode determination is based on a Green's function procedure due to Weyl and others for the explicit evaluation of complete sets of eigenfunctions of differential operators. This formalism was employed to obtain a variety of complete sets of orthogonal functions applicable for alternative representations of electromagnetic fields in spherical regions (EM-29). Each representation has different convergence properties depending upon the parameters involved. Particular applications were made to piecewise constant spherically-stratified regions in order to illustrate the phenomenon of surface waves as opposed to the so-called residue waves. The formalism has also been employed to study the mode structure in general cylindrical regions including open and closed regions with variable isotropic and anisotropic media (EM-69 and current work). The mode solution of electromagnetic field problems in such regions requires the knowledge of the appropriate vector eigenfunctions $E_a(x, y)$ and $H_a(x, y)$ distinguishing the cross-sectional or transverse electric and magnetic field behavior of the a-th mode. In general these possess the vector orthogonality property

$$\iint (E_a \cdot H_b \times z_o + H_a \cdot z_o \times E_b)\, ds = \delta_{ab}$$

over the cross-section transverse to the z-axis of the regions. When suitable symmetry obtains this orthogonality property reduces to

$$\iint E_a \cdot H_b \times z_o\, ds = \delta_{ab},$$

and, in conventional cases, this condition reduces to

$$\iint E_a \cdot E_b\, ds = \delta_{ab} \quad \text{or} \quad \iint H_a \cdot H_b\, ds = \delta_{ab}.$$

Generally the complete set of modes involves both a discrete and a continuous spectrum. These papers present the explicit modes E_a, H_a for a number of cases which illustrate the relative quantitative importance of the ordinary and surface waves of the discrete spectrum (the surface waves are characterized by an exponentially decreasing behavior outside of dielectric regions in which they propagate) and the radiated waves of the continuous spectrum. The contribution of the continuous spectrum to the field has been shown to be negligible in propagation over a spherical highly conducting Earth (EM-28).

The eigenvalue problem arising in the problem of propagation beyond the horizon using the flat-earth modified-index model with a bilinear profile for the index of refraction has been investigated by others in great detail. However, there is some question as to whether the failure of theory to agree with experiment in so far as fields deep in the diffraction zone are concerned is due to the assumption of a flat earth or to the bilinear approximation of the modified index. The claim has been made that if a large number of modes is taken into

---

[2] See also the discussion of the inverse propagation problem in a later section.

account then this bilinear, flat earth theory is adequate. However, the agreement between theory and experiment so obtained may be fortuitous and due to the particular model. It has been deemed desirable to consider directly the spherical earth and a smooth profile for the index of refraction. As a consequence one must determine the eigenvalues and eigenfunctions for this case and we have undertaken to do so, starting with the simpler case of a homogeneous atmosphere. While the eigenvalues of this problem are given correctly for the lower-order modes in standard references, the values for the higher-order modes are obtained from an asymptotic form of the Hankel functions whose numerical accuracy is open to question. This case and the more realistic case of the inhomogeneous atmosphere are currently being explored.

In view of the difficulties in calculating the higher-order eigenvalues we are also exploring through idealized but somewhat realistic models the maximum fields that can be expected in the diffraction zone in order to determine whether the refractive effect of the atmosphere and diffraction by the earth could possibly account for the high fields obtained experimentally. Many investigators are convinced that full explanation of the large, fluctuating diffraction fields obtained experimentally at distances of several hundred miles from the transmitter and at frequencies well above those reflected by the ionosphere is the scattering of radiation from large irregular air masses. Several *ad hoc* theories on the effect of such scatterers have been developed by other workers. We are therefore attempting a fundamental approach to the problem of scattering by a distribution of point scatterers contained in a finite volume of an otherwise homogeneous medium. At the present time our objective is to extend the theory of Foldy, developed by him for scalar waves, to the scattering by an assemblage of Hertzian oscillators. The volume is assumed to be excited by a plane wave and we seek the coherently scattered field. The particles are assumed to have the same spatial probability distribution function and to be mutually independent, and the final field is obtained as an average over all configurations or ensembles. Using the Foldy formalism we have succeeded in reducing the problem of obtaining the average field to one of wave propagation in a continuous medium; in particular, the well-known Lorentz-Lorenz formula for the index of refraction in dispersive media has been obtained. There remains the problem of justifying the fundamental approximation employed by Foldy.

Some attention is being given by electromagnetic research men generally to the important question of whether by choice of a proper "earth" and/or by a proper distribution of sources one can concentrate radiation along the earth at the expense of the field radiated in other directions and thereby increase the effective range of propagation. Application of such theory could be made also in the design of special trapped wave antennas and coated transmission lines (Gobau line).

We have investigated the effect of a dielectric slab over an imperfect earth (EM-65), the case of a perfect earth having been investigated by others outside our group. The source is a vertical dipole. The results are asymptotic formulas for the components of the electromagnetic field at points within and just above the slab but not too near the dipole. This work also yields the dependence of the propagation constants of the modes in the slab on the slab thickness, ground conductivity, and other factors. The present phase of the work is being directed toward obtaining an extension of the asymptotic formulas which will enlarge their domain of validity and in particular will permit a quantitative discussion of how the surface-wave modes cut off. This extension is based on certain results of van der Waerden by which he has generalized the ordinary steepest descent formulas for the asymptotic representation of integrals. In connection with this problem we call attention to the general theory of mode representations mentioned earlier in this section.

The preceding investigation considers the effect of a coating layer but the possibility of exciting surface waves or at least increasing the radiation in the horizontal direction by a proper choice of sources has also been investigated. We have considered (EM-70) the reflection and refraction of an arbitrary electromagnetic wave by a dielectric slab separating two different media. Explicit expressions are obtained for the radiation field in terms of the incident radiation field. These expressions are valid when the point of observation is not too near the plane interface. The method is extended to cover the case of one surface deviating slightly from a plane.

Another study considers an infinite anisotropic slab of thickness d backed by a perfectly conducting metal sheet. For simplicity we assume that the dielectric constants of the slab in the three principal directions are $\epsilon_o, \epsilon_o, \epsilon_3$, where ϵ_o is the dielectric constant of free space. We show that this structure will support surface waves in certain directions only. These directions will be in a dihedral angle perpendicular to that axis of the slab for which the dielectric constant is ϵ_3. We now consider the far field produced by a point dipole placed above the slab. In directions outside of this dihedral angle the field will be that of a space wave which varies inversely as the distance from the source. In the other directions the field will be made up of a space wave and a surface wave which decays exponentially with distance from the slab while on the slab it varies inversely as the square root of the distance from the source.

A current investigation has extended results obtained by Cullen on the excitation of surface waves. Cullen considered a horizontal line magnetic source located at height h above a corrugated surface and showed that there exists a height h_o such that the leading term in the asymptotic expansion of the radiation field vanishes at the surface; he also found a height h_1 at which the ratio of the energy per unit area in the surface wave to the energy per unit area in the radiation field is a maxi-

mum. We have considered both two- and three-dimensional cases; the corrugated surface is replaced by a dielectric medium of finite or infinite thickness with dielectric constant a function of height. In the two-dimensional case our source is a horizontal magnetic line source and in the three-dimensional case it is a horizontal circular ring of magnetic dipoles. The first result of Cullen has been obtained for these two cases. Cullen's second result is now being investigated for our conditions.

A number of special studies in propagation will be mentioned briefly merely to call the attention of interested readers to their existence. We have treated (EM-76) the two-dimensional problem of the propagation of an acoustic wave in the xy-plane above the line $y = 0$ wherein the index of refraction is of the form $f(x) + g(y)$ with $g(y)$ monotonically increasing to infinity with y. Both the pulse and steady-state solutions are obtained. The results include the field in part of the shadow region; the method is an extension of that used in EM-64 to be discussed under diffraction.

Numerous investigators have attempted the problem of reflection and transmission over a rough sea. Their results must be regarded as tentative because no one is quite sure as to what the most useful idealization of the problem should be. We have completed two papers on the subject (EM-40, 58) but can make no affirmation as to their usefulness. Currently we are attempting to analyze the differences in air-to-air communication over land as against over the sea.

We have considered propagation through a cold front (EM-23), which idealized amounts to two semi-infinite homogeneous dielectric media separated by an oblique plane boundary, all over a perfectly conducting flat earth. The resulting fields are expanded in powers of the difference between the dielectric constants of the two semi-infinite media. The terms retained in the solution yield the geometrical optics field plus diffraction effects from the corner. Another isolated propagation study is that of propagation from land to sea in a homogeneous atmosphere (EM-46); the two-part boundary condition is handled by the Wiener-Hopf integral equation method.

The Inverse Propagation Problem and the Synthesis Problem

There are two closely-related problems which concern the differential equation

$$\frac{d^2 u(k, x)}{dx^2} + \left[k^2 - V(x) \right] u(k, x) = 0$$

taken over the infinite interval $-\infty < x < \infty$. This equation and others reducible to this one are involved in tropospheric and ionospheric propagation with no magnetic field, the theory of waveguides with variable cross-section, transmission lines with variable inductance or capacitance, propagation through an infinite slab with dielectric variable in the direction perpendicular

to the faces of the slab, and other physical phenomena. The function $V(x)$ characterizes the inhomogeneous structure in each case. The inverse propagation problem would then be to determine $V(x)$, the index of refraction, from a knowledge of the asymptotic behavior of the solution at large distances. This inverse problem has been solved on the semi-infinite interval by Gelfand and Levitan. We have solved the problem in an infinite interval provided $V(x) \equiv 0$ for x less than some number a (EM-74).

Solutions of the above equation exist of the form

$$e^{ikx} + B(k)e^{-ikx} \quad \text{for} \quad x \rightarrow -\infty.$$

Here $B(k)$ is the reflection coefficient as a function of the wave number. Given $B(k)$ as an analytic function of complex k [it is analytic under proper assumptions on $V(x)$] for waves vertically incident on the troposphere or ionosphere, it is possible to determine the index $V(x)$. Also, if $B(k)$ is known as a function of the angle of incidence of plane waves, all at one frequency, (in which case k involves the angle of incidence α) we can calculate $V(x)$.

Where the medium or circuit is at our disposal and can be constructed, the same theory solves a number of what are called synthesis problems. Thus, if the reflection coefficient of a layer is specified as a function of the angle of incidence of plane waves incident on the layer we can determine what the variable index of refraction of the layer must be and construct one accordingly. Also, if the reflection coefficient is specified to be less than a given quantity for all frequencies of a specified band, the index of the layer can be determined. Further, if we are permitted to approximate the desired reflection coefficients by rational functions then $V(x)$ can be obtained in closed form. This theory is applicable to the design of radomes. The same synthesis problem occurs in non-homogeneous transmission line and waveguide design and in the construction of filter networks.

This work is being continued with particular emphasis on the synthesis problem. It is desirable for example, to have some criteria applicable to the reflection coefficient which would force $V(x)$ to vanish outside a finite interval or to satisfy other conditions. These questions arise from the problem of the physical realizibility of $V(x)$. The objective is broadband synthesis work.

Generalizations of the above results are being pursued in two different directions. The first of these is to consider operators other than the one above and which might apply to problems connected with electric circuits consisting of discrete elements. The second direction is generalization to more dimensions, e.g., to the equation $\Delta u + (k^2 - V(x, y, z)) u = 0$. Typical of the questions one seeks to answer in this connection is the following. Suppose that an amplitude function $A(d_1, d_2, k)$ is specified where d_1 is the direction of an incident wave, d_2, of a transmitted wave, and k, the wave number, can one find an index $n(x, y, z)$ such that a plane wave incident in the direction d_1 will have the far

zone amplitude A in the direction d_2? This problem arises in the design of wide-angle, variable index microwave lenses. Some results applicable to the theory of quantum fields (our report CX-18 not listed in the bibliography) have been obtained for the related problem of determining an unknown linear Hermitian operator from a known operator and a spectral measure function associated with the two operators. In the special cases already studied a knowledge of the spectral measure function is equivalent to a knowledge of the reflection characteristics of the eigenfunctions of the unknown operator.

A more limited investigation in the domain of the inverse propagation problem and applicable to the ionosphere has been completed. A major problem in ionospheric research is that of deducing information on the structure of the ionosphere from soundings by electromagnetic waves. If one neglects collisions and the earth's magnetic field, the electron density can be calculated from the time of travel of a pulse sent up to the ionosphere and reflected by it, provided this time is known as a function of frequency from zero to some frequency and, what is more important, provided the electron density increases monotonically with height. In the case of multiple layers (those with more than one
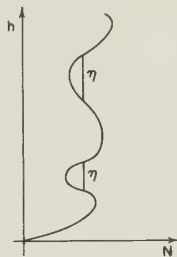
Fig. 5

maximum) the method fails. (Fig. 5). We can ask the question: Suppose we had more information than soundings from the ground, how could we utilize them? Continuing to treat propagation in the ionosphere from the group velocity point of view, we could postulate that we have a station above the ionosphere, and that we can make soundings from below, above, and through the ionosphere. The last type of information, the time of transit of a pulse through a valley (a region between maximum ion densities) as a function of frequency (from penetration frequency to infinity), permits us to calculate the distribution in the valley though not uniquely. All we can hope to obtain is the width of the valley at any density, $\eta(N)$. For more complicated distributions we can obtain the cumulative width of the valleys, $\Sigma\eta(N)$. Since the mathematical procedure involves taking a Mellin transform and later an inverse Mellin transform it is applicable, for the time being, to transit times which are analytic functions of frequency; a suitable approximation procedure would have to be worked out to handle numerical data.

Since transit time is equal to the derivative of the phase of the transmission coefficient with respect to frequency, the problem is closely related to that of calculating medium properties from the scattering matrix. Because the above method uses the group velocity only, we cannot obtain the ion distribution uniquely, but the procedure is simpler and yields such information as the depths of valleys, curvature at the bottom, and other quantities.

DIFFRACTION THEORY

Under this subject we shall consider problems generally involving homogeneous media with obstacles represented by sharp discontinuities in the medium. These obstacles may be pure dielectrics or possess conductivity.

We shall begin with a discussion of a method for obtaining the far field scattered by a combination of obstacles when the far field scattered by the individual obstacles is known. Let us consider two obstacles L and R of arbitrary shape and size but separated by a distance large compared to their sizes. We may regard the field at any arbitrary point P as due to the incident field, the (total) field scattered by L and the (total) field scattered by R. If P is sufficiently far from L, the field scattered by L is of the form

$$U_L(P) = C_L \frac{e^{ik\rho}}{\sqrt{k\rho}} F_L(\theta), \tag{14}$$

where ρ and θ are the polar co-ordinates of P with respect to some arbitrary origin in L since U_L is a far field, and $F_L(\theta)$ is a function expressing the scattering amplitudes in various directions. Similarly there is an $F_R(\theta)$. The scattering amplitude of the combination is $F_L(\theta) + F_R(\theta)$ (where the origin is arbitrary since we are concerned with far fields). Let F_{LO} denote the scattering amplitude of L in the absence of R and F_{RO}, the scattering amplitude of R in the absence of L. We first obtain $F_L + F_R$ in terms of F_{LO} and F_{RO}.

Let us use the notation $F(\theta, \theta_o)$ to indicate the scattered field in the direction θ due to an incident wave having direction θ_o. If now P is on L, by (14)

$$U_R(P) = \sqrt{\frac{2}{\pi k}} e^{-i\pi/4} \frac{e^{ikd}}{\sqrt{kd}} e^{-ikx_P} F_R(\pi, \theta_o)$$
$$= \gamma\, e^{-ikx_P} F_R(\pi, \theta_o), \tag{15}$$

because with respect to the origin in R, the direction of P is approximately π, d is the distance between the origins in L and R and x_P is the x-co-ordinate of P
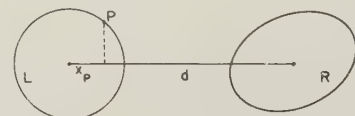
Fig. 6

with respect to the origin in L (Fig. 6). That is, the effect of R on L is that of a plane wave with unknown constant amplitude $F_R(\pi, \theta_o)$. The effect of L at a

general point P is the scattering due to L alone of the incident wave and the scattering by L of the plane wave incident on L from R. Hence we may write for the far field amplitudes

$$F_L(\theta) = F_L(\theta, \theta_o) = F_{LO}(\theta, \theta_o) + \gamma F_R(\pi, \theta_o)F_{LO}(\theta, \pi). \quad (16)$$

Similarly,

$$F_R(\theta) = F_R(\theta, \theta_o)$$
$$= F_{RO}(\theta, \theta_o) + \gamma F_L(0, \theta_o)F_{RO}(\theta, 0). \quad (17)$$

If we set $\theta = 0$ in the first equation, set $\theta = \pi$ in the second equation, and remember that F_{LO} and F_{RO} are supposed known, we obtain two linear equations in the two unknowns, $F_R(\pi, \theta_o)$ and $F_L(0, \theta_o)$. Hence $F_L(\theta)$ and $F_R(\theta)$ are determined.

It is now possible to obtain the approximate total field of the combination in terms of the assumed known fields of the separate obstacles. Again using the subscript O to denote the field scattered by an isolated obstacle we may say that

$$U(P, \theta_o) = U_{\text{inc}}(P) + U_{LO}(P, \theta_o)$$
$$+ \gamma F_R(\pi, \theta_o)U_{LO}(P, \pi) + U_{RO}(P, \theta_o)$$
$$+ \gamma F_L(0, \theta_o)U_{RO}(P, 0). \quad (18)$$

In words, the total field at P due to a wave incident from the direction θ_o is the incident field, plus the field scattered by L alone due to the incident field, plus the field scattered by L due to the plane wave incident upon it from R, which has the direction π from L, plus the field scattered by R alone due to incident field of direction θ_o, plus the field scattered by R due to the plane wave incident upon it from L, which has the direction O from R. Since γ is known and the quantities $F_R(\pi, \theta_o)$ and $F_L(0, \theta_o)$ were obtained previously the total far field is known. The third and fifth terms in (18) may be regarded as the correction terms to single scattering theory.

The coefficients of the correction terms are quotients with denominators of the form $1 + 0\left(\dfrac{1}{\sqrt{kd}}\right)$. If the denominators are simplified by ignoring the higher order terms we obtain a general formula for first order multiple scattering which may be adequate in some problems. However, as noted below, there are cases where the higher-order terms should not be neglected.

The above method holds even if the size of each body is comparable to or larger than the spacing provided each body subtends a small angle when viewed from any other body of the combination of obstacles, for example, two half-planes forming a slit. In this case the respective subtended angles are 0 even though the bodies are infinite in size. In so far as the slit problem is concerned, strictly each body is excited by a line source rather than a plane wave. However there is a little known reciprocity theorem which states that the far field in the direction θ of a line source is expressible in terms of the near field of a plane wave incident from the direction θ and evaluated

at the line source. When the line source is itself remote we can replace the near field of the plane wave by the far field of the plane wave, thus obtaining a further simplification because the coupling effect still involves only the response of cylinders to plane waves.[3]

The above method can be applied to n bodies with the modification that if the objects are all different then one must then solve $n(n - 1)$ simultaneous algebraic equations. However in the case of a grating, where the obstacles are congruent and in one line, only two algebraic equations must be solved. The problem of diffraction by a grating has, of course, been treated by numerous authors under the restrictions that either the elements are infinitesimal cylinders, the ratio of spacing to wavelength is small, the interaction of the cylinders is neglected, or the cylinders are taken to be circular. The above method allows the grating elements to be arbitrary congruent cylinders of arbitrary size; however, it assumes that the largest diameter of an element and the wavelength are small compared to the spacing. It supplies a relation between the spectral amplitudes for the grating and the far field amplitudes of an isolated element. If the higher order terms in the denominator of the correction term are neglected then the result is good only on the assumption that there is no spectrum in the plane of the grating.

This so-called resonance case results under normal incidence when the wavelength equals the spacing. In this case the procedure of simplifying the denominator will not work because the sum of the terms to be neglected become infinite. Thus the first-order multiple-scattering result is inaccurate. When the full denominator in the correction term is retained the resulting expression is indeterminate. We have however evaluated the limit as resonance is approached. The results involve certain second- and third-order determinants formed from the array of single scattered amplitudes. The vanishing or nonvanishing of the determinants is critical for the calculation. Detailed consideration of the result shows that, in contrast to the nonresonance case, interaction can never be neglected near resonance. The intensity in each spectrum for wavelengths near resonance can be markedly different from that for other wavelengths, and our formulas exhibit this feature as well as a dependence on the resonance of the single scatterer. We believe that these results bear directly on the well-known problem of grating anomalies, which has been discussed analytically by Artman for the case of a reflecting grating. But the results also indicate that anomalies are conceivably not a necessary consequence of resonance in the case of conducting cylinders; the shape of the cylinders may be equally significant.

Before leaving the subject of the above method, which has been discussed here primarily in connection with gratings and slits to which it has been successfully applied, we should like to mention that the method seems

---

[3] The slit problem proper is discussed below.

applicable to such problems as diffraction by a pair of parallel, long but finite plates and by a finite cylinder.

Turning to specific classes of diffraction problems we mention first some other work on gratings. We have treated the diffraction of a plane wave by a semi-infinite array of parallel, equidistant, congruent conducting cylinders The object of the work is to compare the results with those for a grating which is infinite in both directions so as to ascertain the end effects and their relative importance. The analysis is materially simplified if we restrict ourselves to cylinders of infinitesimal cross section. The currents A_n which are induced in the n-th cylinder ($n = 0, 1, \cdots$) are then determined by an infinite system of algebraic equations. Considering interaction of each wire with only its N nearest neighbors, we solved this system by very simple considerations involving the generating function $A(z) = \sum\limits_{o}^{\infty} A_n z^n$. If total interaction is considered the system of equations can be solved by the use of the *discrete* Wiener-Hopf theory, such as is employed in Wiener's theory of discrete prediction. A method is now being studied for the evaluation of the resultant cumbersome expressions.

In earlier papers devoted to the grating problem (EM-26, 34, 39, 54, 58, 59) we approached the grating as a set of half-cylinders (also hemispheres) equally and then randomly spaced on a conducting plane. In these papers the attempt to include multiple scattering, that is, the interaction of the elements on each other, was handled by a purely algebraic summation of the fields scattered by the individual elements when illuminated by the incident plane wave and by the first, second- and higher-order fields scattered by the other elements. However, the resulting expressions were clumsy and were useful only for special cases such as a grating of two or three elements and for particular directions of the incident wave such as grazing incidence. In two other papers on the grating (EM-18, 41) the variational method was applied to a plane grating consisting of equally spaced cylindrical wires illuminated by a plane wave incident at right angles to the grating element but not necessarily to the grating plane. The amplitudes of the spectra of various orders were calculated by using as a trial current distribution that induced in a single wire by the plane wave. Interaction effects are included.

The classical problem of diffraction by an aperture in a screen and the complementary problem of the disc has attracted our group as it has others. Interest in the problem is both mathematical and practical. On the mathematical side it is one of the few problems that has been solved exactly in both the scalar case (Bouwkamp) and the vector case (Meixner and Andrejewski). One can therefore study electromagnetic techniques by testing them on this problem or use the aperture problem as a model for the investigation of more complex problems. On the practical side this theory has been applied to irises and slots in waveguide and antenna structures.

Current research on this problem including our own is concerned with approximate solutions because the exact solutions are laborious to compute. In one of our early papers on this subject (EM-19) we compared the Kirchhoff method for treating acoustic diffraction by a disc with Bouwkamp's exact solution. The comparison established the fact that the Kirchhoff method is more accurate than commonly supposed.

The circular aperture problem for a normally incident plane wave was treated variationally by Levine and Schwinger. The exact diffracted field is obtainable from a knowledge of the correct field in the aperture. To obtain the latter field these authors used the static field as a trial function. The essential difficulty in their approach is the solution of an infinite system of linear equations, the unknowns of which are the coefficients of an expansion of the true field in the aperture in terms of a properly chosen set of functions.

We have shown (EM-32) that the matrix of this system of equations is of a comparatively simple nature. Its investigation provides a rigorous proof of the convergence of the perturbation procedure for sufficiently small values of $ka (k = 2\pi/\lambda, a$ = radius of the aperture). Bouwkamp proposed a change of the set of functions in terms of which the field in the aperture is expanded. Using his set, the matrix of the infinite system becomes the identity for $ka = 0$ (static case) and tends towards a matrix M_∞ which has a formal inverse for $ka \rightarrow \infty$ (optical case). Also, the new system can be shown to be equivalent with a Fredholm equation of the second kind which has a unique solution for all real values of ka. Attempts are being made to utilize these facts for a better analysis of the behavior of the diffracted field both for large ka and for an investigation of the transmission coefficient as an analytic function of ka.

The transition from large ka to $ka = \infty$ in which the aperture field for large ka appears as a perturbation on the $ka = \infty$ or optical case presents special difficulties. If $ka \rightarrow \infty$ the Fredholm integral equation of the second kind degenerates into one of the first kind. For the limiting case the equation can be solved explicitly, giving the optical solution. However the transition from large ka to $ka = \infty$ involves complicated asymptotic forms of the kernel of the Fredholm equation and these are being studied.

The exact series representations of the scalar and vector fields diffracted by a circular aperture, which involve spheroidal functions, lend themselves readily to calculation at long wavelengths but prove unsuitable at short wavelengths where even verification of the geometrical optics limit is analytically difficult. Hence we are currently considering a plane wave normally incident on the circular aperture in soft or hard screens ($\Psi = 0$ or $\partial_n \Psi = 0$ on the screen) and are seeking the explicit lowest-order correction, in reciprocal powers of ka, to the geometrical optics transmission cross section. The method is the variational one applied to an integral

equation formulation of these problems. Thus, for the soft screen, an integral equation describing the distribution of $\partial_n\Psi$ on the shadow face affords a convenient basis. By appropriate modification of the Green's function occurring in the integral equation, the high frequency behavior of $\partial_n\Psi$ can be obtained, and utilized for the calculation of the transmission cross section. The first approximation to $\partial_n\Psi$ is that pertaining to a local infinite straight edge at each direction tangent to the aperture rim, as proposed by Braunbek on physical grounds. In the next approximation, a supplement to the Braunbek distribution is again associated with a straight edge configuration, where however, the excitation arises from a (secondary) plane wave at grazing incidence. The grazing wave describes an interaction between points at (roughly) opposite ends of an aperture diameter. It is noteworthy that even a first correction to the geometrical cross section obtained in this way proves reliable at moderate frequencies, a feature well known in other instances of asymptotic expansions. This circumstance contrasts markedly with the behavior at low frequencies, where successive terms in a power series expansion do not significantly extend the qualitative or quantitative aspects of the cross section.

We are also considering currently diffraction by apertures of arbitrary shape. We seek to carry the analysis as far as possible without specification of shape and to examine the relationship between the vector and scalar cases. By an analysis too detailed to present here we have found that the determination of the solution of the vector diffraction problem in terms of the solution of the scalar problem depends upon the solution of an integro-differential equation for a function, $c(s)$, which expresses the coupling between the electric and magnetic fields determined by the edge. The kernel of this equation depends only upon the geometry of the aperture.

The present formulation makes it possible to examine the existence question for the problem of vector diffraction by a plane screen with an arbitrary smooth rim, under the assumption of the existence of scalar Green's and Neumann's functions which are $O(\rho^{1/2})$ at the rim. An investigation of these questions was therefore commenced, with the following preliminary results: (1) We can show rigorously that any scalar wave function which tends to zero uniformly as the screen (including the edge) is approached has a Taylor's development in powers of $\rho^{1/2}\sin\phi/2$. If the edge is analytic we have an infinite power series. This result was obtained by the introduction of local parabolic co-ordinates, $\xi = \sqrt{\rho}\cos\phi/2$, $\eta = \sqrt{d}\sin\phi/2$. These act as uniformizing variables and permit the application of a generalized reflection principle. The result makes possible the proof of a simple uniqueness theorem for bodies with edges. (2) We have set up the problem of determining the scalar Green's function in terms of an integral equation of the first kind, by a representation of the scattered field as a simple layer. Then, conformally mapping the

domain of the integration into the interior of a circle and using the known static Green's function for a circular disc, we converted the equation into an integral equation of the second kind. The properties of the resulting kernel are now to be investigated, with a view to applying the Fredholm theory. A similar attack in terms of the static solution for the disc is also being examined.

The two-dimensional aperture problem, that is, the diffraction of a plane wave by a slit in an infinite conducting plane, has also been treated. Mention of one such investigation was made earlier in this section and is here elaborated. A rigorous solution in terms of Mathieu functions has been given by Morse and Rubinstein and by Skavlem, and approximate solutions have been given by Sommerfeld, Bouwkamp, and Groschwitz and Hönl. If d is the *half-slit* width, λ is the wavelength, and $k = 2\pi/\lambda$, then we can say that the rigorous solution is useful for $kd < 10$, while the approximate solution mentioned above is useful only for much smaller kd. On the other hand Schwarzschild has demonstrated a series of successful approximations which converges for all kd, but is especially useful in the limit $kd \to \infty$. Schwarzschild evaluated the leading term, which is useful only in the optical limit and which represents the half-planes as scattering independently. The present work bridges the gap in the range of kd covered by existing solutions and obtains the form of the correction term[4] in the expansion of the solution for large kd. It proved possible to show *a priori* that the correction term could be represented as a superposition of the fields scattered by each half-plane separately, when each half-plane is under the influence of a line source of a suitable complex constant amplitude. The necessary amplitude could be determined from a pair of algebraic equations. For the case of a normally incident plane wave, the resulting far field angular pattern was calculated at 10-degree intervals and for 12 values of kd in the range $3 \leq kd \leq 8$. These results were compared with the results which we calculated from the exact Mathieu function solution. It was found that our interaction term correctly supplies an essential correction to Schwarzschild's results. One may conclude that for $3 \leq kd \leq 8$, and *a fortiori* for larger kd, the present elementary solution may be employed with confidence (EM-75).

The problem of diffraction of electromagnetic waves by an infinite slit in a perfect conductor at the interface between two *different* dielectric media has also been solved by expanding the fields in each medium in terms of appropriate Mathieu functions (EM-68). By matching these expansions across the slit an infinite set of linear equations is obtained for the coefficients in the expansion of the diffracted fields. The solution is put in such a form that numerical results can be readily obtained.

---

[4] Reference should be made to the discussion of the method described at the beginning of this section.

Of considerable mathematical interest, especially in connection with diffraction problems, is the behavior of electromagnetic fields at edges or corners. In one paper (EM-71) we investigated the effect of a local discontinuity of dielectric constant upon the singularity in field strength which occurs in the neighborhood of a geometrical singularity of conductor. The work is confined to the electrostatic case. We consider first the case of a wedge-shaped conductor placed between two different dielectrics so that the plane interface passes through the vertex of the wedge. For this problem the dependence of the singularity on the various parameters is studied in detail and, in addition, the Green's function is derived. It is found that when the wedge is symmetric with respect to the interface the singularity is unaffected by the presence of a discontinuity. Also, if the greater angular region has the greater dielectric constant the singularity is increased and vice versa. It is also noted that, in the symmetric case, the Green's function can be expressed, by the method of images, in terms of the Green's function for a conducting wedge in free space. The second of these results is then extended to the case of an arbitrary symmetric distribution of conducting material. The extension of the first result is then deduced.

In another paper (EM-72) the behavior of an electromagnetic field in the neighborhood of the common edge of angular dielectric or conducting regions is determined from the condition that the energy density must be integrable over any finite domain (the so-called edge condition). Two cases are treated in detail, namely, (1) a region consisting of a conducting wedge and two different dielectric wedges with a common edge, and (2) a region consisting of two different dielectric wedges with a common edge. It is also shown that near such edges electrostatic and magnetostatic fields will exhibit the same behavior as the electromagnetic field.

The singularity at the corner of a plane sector due to a dipole field has been approached through the use of the sphero-conal co-ordinate system of which the sector is a co-ordinate surface. This co-ordinate system is made up of two families of right elliptic cones whose axes are orthogonal at the common apex and a family of concentric spheres about this point. The co-ordinate system is orthogonal; also, there are solutions of the scalar wave equation in the form of a product of three factors each of which is a solution of an ordinary differential equation. The boundary conditions give eigenvalue problems for each of these equations. One equation has as its solution the spherical Bessel functions. The other two equations are Lamé equations. The Green's function can then be obtained without difficulty. Basically then, the problem has been reduced to finding the eigenvalues and eigenfunctions of Lamé's equation. The procedure for studying this problem is analogous to that used in the theory of Mathieù functions and spheroidal functions. Expansions of the Lamé functions have been

obtained in terms of half-angle trigonometric functions and regular (at ± 1) associated Legendre functions. In general, three-term recurrence formulas for the coefficients of the series are obtained in both cases. The simultaneous solution of these recurrence equations is being studied. One set of eigenvalues determines the behavior of the solution near the sharp corner of the sector.

We mention in passing that like other research people we have studied the diffraction of pulses, point sources, and line sources by discs, wedges, cylindrical surfaces of various cross sections, the cone, the sphere, thin shells, truncated cones, paraboloids of revolution, wedges with rounded corners, the corner reflector, and other shapes (EM-21, 36, 43, 45, 52, 53, 64, 66, 172-5). Reference is also made to some diffraction studies noted previously. We recognize, of course, that these problems have been worked on by numerous other investigators. However, in each of the above-mentioned investigations there is some aspect, methodology or a more useful representation of the solution, which we believe to be new.

An example taken from current work is the old problem of the scattering of an incident plane wave by a conducting sphere. We are concerned with obtaining a rapidly converging asymptotic form for large ka of the wave scattered by a perfectly conducting sphere of radius a. A solution for the total field expressed in terms of the radial eigenfunctions has been obtained. The eigenvalues for these functions are complex due to the non-Hermitian boundary condition at infinity. These eigenvalues may be obtained by a saddle point approximation, but the initial values occur where the saddle point method breaks down. For these values, an expansion near the turning point is made in terms of Hankel functions of order one-third. The behavior of the Hankel function for three different radii, the radius of the sphere, the distance to the source, and the distance to the observer is needed. To facilitate obtaining this information, the behavior of the Hankel function is discussed in terms of their saddle points using the Sommerfeld integral representation. A method for constructing the paths of steepest descent is found that is valid for all saddle points and that reduces to the "Debye contours" when the index is real.

Finally, a review paper (EM-50) which discusses critically recent work in diffraction theory with special reference to the aperture problem should be noted.[5]

The Theory of Antennas

A large number of the research efforts described above in connection with quasi-optical field theory, propagation, and diffraction bear directly on the theory of antennas. For example, the determination of fields diffracted by parabolic cylinders or paraboloids, the theory of surface waves, the calculation of the field on a

[5] A broader review of current work in diffraction theory is given by C. J. Bouwkamp, "Diffraction Theory," Reports on Progress in Phys. vol. 17, pp. 35–100; 1954.

caustic of a parabolic cylinder, and the work on the asymptotic field diffracted by an idealized lens system are applicable to antenna systems and to microwave dielectric lenses. However, a few of our papers were undertaken with the antenna problem specifically in mind and we shall therefore discuss them here.

Several papers have been devoted to the broad problem of determining the possible radiation patterns and the relationships of these patterns to sources. We treated first the scalar problem and obtained (EM-62) a necessary and sufficient condition that a far field angular pattern must satisfy in order that it be the radiation pattern of some distribution of sources enclosed in a finite volume. From this theorem it follows, in particular, that no far field pattern can be identically zero in any angular sector. The corresponding vector problem has just been treated. An electromagnetic radiation pattern is defined as a tangential vector field on the unit sphere; this field describes the asymptotic behavior at infinity of the full electromagnetic field. The square of the modulus of this tangential field determines how the radiated energy depends upon direction. As in the scalar case, a necessary and sufficient condition that a tangential vector field on the sphere be a radiation pattern is derived.

After discussing the properties of radiation patterns, the problem of generating these patterns by sources is studied. Each pattern belongs to some electromagnetic field existing outside the sphere and, hence, it is possible to generate this pattern by sources contained within the sphere. However these sources are not uniquely determined by the external field and we therefore consider the question of the most efficient distribution of sources. The concept of efficiency is first defined for a finite array of dipoles located at fixed points. If we assume that these dipoles radiate incoherently the total radiation is easily determined and is, apart from a numerical factor, the sum of the squares of the amplitudes of the dipole fields. However, if the dipoles radiate coherently, the fields of the individual dipoles interact and the radiated energy may be made larger or smaller than that obtained in the incoherent case. The efficiency of the distribution of sources is then defined as the ratio of the coherent to the incoherent radiation.

The analysis of efficiency is pursued by using the algebraic theory of quadratic forms. For homogeneous media the coefficients of the forms may be obtained directly. To treat the efficiency of a continuous distribution of dipoles (currents) contained on a line, surface, or region, the algebraic quadratic forms are replaced by quadratic integral forms and the theory of linear symmetric integral equations gives a complete description of the efficiency.

Currently we are investigating the field which is diffracted by an infinite parabolic cylinder and which arises from a line source parallel to one of the generators but not coinciding with the focal line of the cylinder. Solution of this problem should lead to a rigorous treatment of the field on the caustic. We are also considering the limits of convergence of the series representing the field diffracted by the cylinder when the source is a plane wave or a line source. Both problems utilize separation of variables on the reduced wave equation in terms of parabolic co-ordinates and lead to Weber-Hermite functions. Special facts about these functions, e.g., addition theorems and their behavior for large values of the variable and of certain parameters, are required and are being investigated.

Typical of the results obtained in our earlier work is the determination of the field radiated by a quarter-wave dipole over a conducting ground of finite radius (EM-19).

The Theory of Waveguides

Much of the general theory discussed under propagation, diffraction, and other headings, for example, the general theory of the representation of fields described in the section on tropospheric propagation, applies to waveguide structures. However, we have obtained some results which are best described separately.

There are some broad principles in electromagnetic theory, such as Babinet's and the reciprocity principles, which tell us how to solve some problems if we know the solution of others. For example, Babinet's principle tells us how to obtain the acoustic field diffracted by a rigid disc from a knowledge of the field diffracted by an aperture in a soft screen. We may call two such problems equivalent, and the principle which relates them may be called an equivalence principle. Another principle whose applicability to electromagnetic problems seems to have passed unnoticed is due to Schwarz. This principle asserts that the unique analytic continuation of a function u across a plane on which $u = 0$ is given by an odd reflection while the unique continuation across a plane on which $u_n = 0$ is given by an even reflection.[6] We shall show by some examples that this principle can serve as an equivalence principle for electromagnetic problems.

Recently in treating a grating problem Heins and Baldwin solved it by reducing it to the problem of a semi-septum in a waveguide and solved the latter problem by the Wiener-Hopf technique. The formulas occurring in the latter solution are reminiscent of those occurring in the bifurcated waveguide problem. We shall indeed show that the septum problem can be solved if the solution of the bifurcation problem is known for an incident plane wave which is symmetric about the bifurcation.

Consider the bifurcation problem in which the boundary condition $u = 0$ is imposed on the bifurcation and $u_n = 0$ is imposed on the walls (Fig. 7). We now continue the $u(x, y)$ of Region I analytically by odd reflection into Region II and, using the fact that in the

---

[6] As a matter of fact, Babinet's principle can be proved by means of the Schwarz reflection principle.
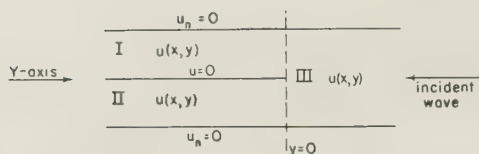
Fig. 7
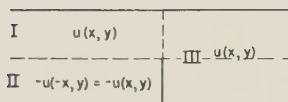
Fig. 8
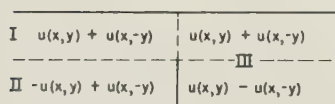
Fig. 9

Fig. 10

Fig. 11
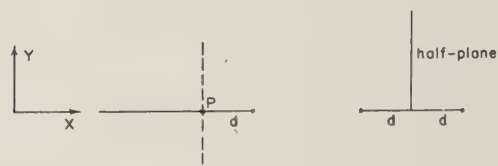
Fig. 12
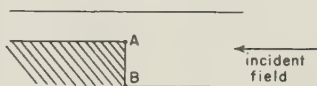
Fig. 13

original solution $u(x, y) = u(-x, y)$, we assign $-u(x, y)$ to Region II. In Region III the original function is retained (Fig. 8). The analytic continuation performed in going from Region I to Region II amounts to going into the second sheet of the multivalued $u(x, y)$ and the bifurcation is a branch line. The function now defined in the entire space is not analytic across the septum and no simple boundary condition holds on that segment. Let us call $v(x, y)$ the function now defined in all regions of the guide with septum. By an even reflection we now construct a function $w(x, y)$ which solves the septum problem. The values of w are shown in the various regions (Fig. 9). On the boundary between Regions I and III w is analytic and $w_n = 0$. On the septum itself $w = 0$ and w_n is discontinuous. Had we chosen an odd reflection we could have satisfied the condition $w_n = 0$ on the septum and w discontinuous there. The above solution corresponds to an incident wave from the left and the right. However the solution due to a single incident wave can then readily be extracted.

The above method is applicable to a variety of problems. From the solution of the problem of a waveguide of arbitrary cross section with a bifurcation which is a plane of symmetry we can obtain the solution of the septum problem in a guide of the same cross section (Fig. 10). The following two eigenvalue problems namely, those of the membranes clamped along the boundary and along the segment shown in Fig. 11, are equivalent. Consider also the half-plane on which two waves one from below and one from above symmetric about the plane $y = 0$ are incident (Fig. 12). By reflection in the plane through the point P and perpendicular

to the half-plane we can solve the problem of a wave incident normally on the inverted T-shaped structure shown in Fig. 12 and coming from $y = \infty$. The solution of the latter problems can in fact be expressed explicitly in terms of Fresnel integrals. Other examples of equivalent problems can readily be given.

In recent years considerable research has been carried out on problems involving discontinuities, such as bifurcations or step discontinuities, in waveguides. For problems involving bifurcations in rectangular and cylindrical waveguide it has been possible to obtain exact solutions by the Wiener-Hopf technique, but for problems involving step discontinuities a different method of attack has in general been employed. A study just completed applies the Wiener-Hopf technique to the problem of step discontinuity in a parallel plate waveguide (Fig. 13).

It has been found possible to write the reflection coefficient for the complete solution as the sum of two parts, one of which can be written explicitly in terms of known functions, whilst the other can be obtained as the solution of an infinite number of linear algebraic equations. The explicit part may be obtained by assuming that the step is absent but that the bifurcating plane remains in a waveguide of the larger width and then superimposing onto the incident field an additional field incident from the region formerly occupied by the step such that the normal derivative of that part of the total solution involving the fundamental mode vanishes on AB. The incident field to be superimposed may be obtained from a knowledge of the exact solution for the radiation from the lower half of the bifurcated region (Fig. 13). The analytic, explicit part of the solution is

seen to be by far the dominant part of the complete solution and is such that a very approximate solution of the set of linear algebraic equations yields a very good value for the total reflection coefficient. Comparison has been made with results[7] obtained by the static method for the case when the regions are such that only the fundamental mode is propagated without attenuation. The results of both methods agree to within less than 1%. The present work has also been extended to cover the case where the larger section can sustain two nonattenuating modes, and values of the reflection coefficients have been computed for various step heights and for different waveguide heights. The method can be extended to cover the case where the step is replaced by a dielectric slab with a perfectly conducting sheet resting on it; this is equivalent to a bifurcated guide when one of the bifurcated regions consists of a medium with dielectric properties differing from that of the main structure. The method is also applicable to radiation from dielectric waveguides and it is possible to obtain an iterative type of solution in terms of the difference between the dielectric constants of the dielectric and of free space. In all problems confined to closed regions it should be possible to recast the solution in such a manner that the dominant part of the solution may be written in terms of known functions whilst the degree of accuracy required for the determination of the other part should be low and yet produce good accuracy for the complete solution.

Among older and more specialized results we might mention a simplification introduced into the calculation of quantities needed in the design of complicated waveguide junctions (TW-16). The quantities needed are the eigenvalues of high-order matrices. Since these matrices generally possess special symmetry properties it is possible to reduce the calculation of eigenvalues to the calculation of eigenvalues of lower-order matrices. This paper shows how the calculation may be so simplified.

Another result in the theory of waveguides (TW-17) concerns the determination of the cutoff frequencies (eigenvalues) of almost circular, cylindrical waveguides. The principal result is the attainment of a better lower bound for the principal eigenvalue. This bound is in the form of an integral over a specially chosen circle related by a parameter to the almost circular boundary being investigated. The result is specialized and detailed but of appreciable theoretical importance.

An old paper (170–1) shows that the cutoff frequencies of a coaxial circularly cylindrical guide approach those of the external cylinder itself as the radius of the inner cylinder approaches zero. Other rather particularized results are indicated in the bibliography.

Finally, in connection with guides we should like to call attention to a series of papers which deal specifically with the helix as a waveguide. The exact solution of the problem of obtaining the electromagnetic fields which

can be guided by a helical wire, even in the absence of an axial electron beam, is very difficult and not likely to be solved. We have therefore formulated and treated several idealizations of the helical waveguide. One idealization (170–2) treated a circularly cylindrical guide with continuous wall structure perfectly conducting in the direction of the true helical wire and perfectly nonconducting in the perpendicular direction. The magnetic field was assumed to be zero outside. Another idealization (170–3, 170–7) considered the same structure and same conditions as to the electric field but called for a magnetic field continuous through the guide wall. In a third idealization (TW-10) the thickness of the wire was specifically taken into account by the introduction of a new nonorthogonal co-ordinate system in which the surface of the wire is a co-ordinate surface. A fourth idealization (TW-11) treated the infinitely thin wire and assumed a sinusoidal current distribution on the wire and propagation constant β in the axial direction. New boundary conditions were introduced to replace the usual condition that the electric field vanish along the wire because the usual condition required an infinite field strength along the wire. The several idealizations are approximations in one respect or another but each gives some information. Moreover, the various approximations serve as a check on each other. The second of the above idealizations is most widely used by other investigators. Our reports on it did furnish information about the higher modes which was not contained in other papers.

In connection with the cold helix we also studied the problem of matching to a coaxial line (TW-12) and the more general problem of matching to an arbitrary given transverse field (TW-13). The orthogonality properties of the modes were investigated in the last mentioned report.

Several of the studies of the helix considered the combination of the helix and an axial electron beam and treated such matters as amplification of the electromagnetic wave fed on to the structure, noise, and the effects of an external magnetic field, particular excitations, electron velocity, and electron current (TW-9, 14, 15, 19).

MATHEMATICAL STUDIES

Progress in the theoretical solution of electromagnetic problems is obviously dependent upon progress in mathematics proper. We therefore investigated a number of mathematical problems lying within our individual spheres of interest and arising in our electromagnetic research.

Much of the work discussed earlier on asymptotic solution of Maxwell's equations and the second-order linear hyperbolic partial differential equation, on spectral representation of various operators in different media and under different types of boundary conditions, and on systems of differential equations contains, of course, mathematical interest over and beyond the

---

[7] N. Marcuvitz, "Wave Guide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y.; 1950.

bearing on electromagnetic problems. Some additional papers on spectral theory (EM-42, BR-4) may be referred to in passing.

A great deal of attention has been given to the mathematical method embodied in what is commonly known as the Wiener-Hopf integral equation technique. One of our basic reports (EM-25) sought to analyze the inner mechanism, characterize its legitimate range of application, and demonstrate its connection with the method of separation of variables. Boundary value problems are generally attacked so that the physical boundary is one entire co-ordinate surface. In this case the usual method of separation of variable can be employed at least in principle. However, there are problems in which the boundary does not consist of one entire co-ordinate surface but, for example, the condition $u = 0$ may have to be satisfied on a half-line and $u_n = 0$ on the other half. We refer to such a problem as a two-part boundary value problem. Moreover, the choice of co-ordinate system may determine whether a given problem is a one-part, two-part, or even three-part boundary value problem. Normally the Wiener-Hopf technique is successful with two-part boundary value problems. However, this report shows that the separation of variables (or eigenfunction) technique when supplemented by function-theoretic techniques (such as are involved in the Wiener-Hopf method) not only solves such problems but is entirely parallel to the Green's function, Wiener-Hopf technique. For example, diffraction by an infinite number of staggered semi-infinite parallel plates, solved in the past by the Wiener-Hopf technique, is solved in this paper by the function-theoretic extension of the method of separation of variables. The paper also shows how the integral equation to which one is led by the use of the Green's function may be solved by the use of transforms other than the Fourier transform used in the Wiener-Hopf method. Thus the infinite ribbon may be treated in polar co-ordinates as a two-part problem and by use of the Mellin transform. (Of course, it can be treated as a one-part problem in elliptic co-ordinates.) A number of old and new problems are used as illustrations in this paper.

It was the above analysis of the domain of application of the Wiener-Hopf technique which made it evident to the author that the technique was applicable to problems such as the charge distribution on a finite conical cup (EM-35; cf. also EM-66). This is a two-part problem and the method of solution is Wiener-Hopf. The solution is potentially useful as a trial solution of an electromagnetic problem and is of intrinsic interest as an electrostatic problem.

The above analysis was pursued further in a later report (BR-11). One-part boundary value problems are in principle treated by separation of variables and the reduction to a Sturm-Liouville problem for the resulting ordinary differential equations. On the other hand two-part boundary value problems are reducible to one-part problems in principle by means of the function theoretic

technique used in the Wiener-Hopf method. Hence it seemed desirable to provide an abstract general theory of two-part problems in terms of the related one-part problems. This has proved possible in the case where the spectrum of the one-part problem is discrete. This analysis is best described in terms of the problem of a bifurcated waveguide in which the medium is stratified in the direction perpendicular to the bifurcation. We suppose that the problem posed by the bifurcated guide with the bifurcation extending to infinity in both directions is solved and that the unbifurcated infinite guide problem is likewise solved. These separate problems are, of course, one-part problems and reduce to one-dimensional Sturm-Liouville problems. The original two-part problem can then be solved in terms of the solutions of the separate one-part problems. The central idea is to construct directly an integral representation of the solution, whose residue expansion has two different forms which correspond to the solution for the two regions. The basic feature of the integral is the presence of certain infinite products in the denominator of the integrand of the form

$$\pi(1 - \frac{\lambda}{\lambda_n})e^{\lambda/\lambda_n}, \qquad\qquad \pi(1 - \frac{\lambda}{\mu_n})e^{\lambda/\mu_n},$$

where λ_n, μ_n are the eigenvalues for the separate problems and λ is the integration variable. These products are a generalization of the Gamma function in the same way that the solution of the Sturm-Liouville problem is a generalization of the sine or cosine functions. The products are related to a factorization of the entire functions in λ which are solutions of the Sturm-Liouville equations in the same manner as the Gamma functions are related to the factorization of $\sin \pi z$ in accordance with the identity $\sin \pi z = \pi/\Gamma(z)\Gamma(1 - z)$.

Another class of mathematical investigations comprises research in special functions. Some new results have been obtained on the subject of periodic solutions of Hill's equation and the special case of Mathieu's equation (BR-1, 6), both of which play a role in diffraction by objects of elliptical shape. These particular investigations concern primarily the analytic character of a transcendental function involving an infinite determinant. The zeros of this function determine the periodic solutions of Hill's equation.

We are presently concerned with the derivation of addition theorems for parabolic cylinder functions. As is well known, the separation of the wave equation $\Delta u + u = 0$ is possible in a few co-ordinate systems only. With the exception of the Cartesian system, all of these co-ordinate systems contain distinguished points or lines. However, the wave equation itself is invariant with respect to all movements of the space. Therefore, if ϕ_v and Ψ_v are two complete sets of solutions of the wave equation which are obtained by separation of variables in two co-ordinate systems with different sets of distinguished points or lines, it must be possible to express the ϕ_v linearly in terms of the Ψ_v and vice versa.

The resulting formulas are known as addition theorems. They are needed for diffraction problems which require, say, a system of polar co-ordinates for their solution but where the source point is not at the origin of the co-ordinate system. Some very general theorems of this type have been given by Friedman and Russek for spherical waves (EM-44). It is the purpose of the present investigation to devise similar results for the functions of the parabolic cylinder and of the paraboloid of revolution. Since these are rather well-investigated classes of functions, no great difficulties are expected. On the other hand, the derivation of the addition theorem seems to be a natural and unifying approach to the theory of these functions.

Some results have been obtained for the parabolic cylinder functions $\omega_n(u, v)$ which involve the Weber-Hermite functions and for wave functions $\phi_n(\rho, \sigma)$ appropriate to the paraboloid of revolution and involving the confluent hypergeometric functions.

Some new integral relations involving spheroidal functions have been obtained from a general integral relation by properly choosing the kernel of the integral (EM-73). Several limiting cases of the relations offer new integral relations among products of associated Legendre functions. Some very general integral expressions involving Bessel function integrands have been derived which satisfy the wave equation and $u_n = 0$ on a screen containing a circular aperture, namely,

$$I_{a,\beta,\sigma}\left(\rho, z, a\right) = \int_0^\infty e^{-z}\sqrt{1 - t^2}\,\frac{J_a(\rho t)J_\beta(at)t^{2\sigma-a-\beta}}{\sqrt{1 - t^2}}\,dt.$$

These integrals are generalizations of integrals obtained by Bouwkamp, Levine, and Schwinger, for the solution of the aperture diffraction problem. Our integrals have been expanded into convergent series, and asymptotic representations for large z and large ka have been derived.

In many physical problems, such as those of wave propagation or quantum mechanics, the solution is expressed as a definite integral and then this integral is evaluated, at least for large values of the parameters, by the methods of stationary phase, saddle point, or of steepest descent. These methods are awkward to apply when the integral has two saddle points close together or even coincident for some values of the parameter. A typical case of this is the asymptotic expansion of $J_n(z)$ when both n and z are large. It is well-known that three different expansions are needed, according as n/z is much less than unity, n/z is much greater than unity, and, the most difficult case, n/z is close to unity. We have devised a method which will given an expansion that is useful no matter what the ratio n/z is. We proceed as follows: Consider the integral

$$\int_c e^{iz(\sin\theta - r\theta)}\,d\theta, \tag{19}$$

which is essentially $H_{rz}^{(1)}(z)$ if a suitable contour is chosen. This integral has saddle points at the values of θ for which $\cos\theta = r$. If $r \pm 1$, there are two values, $\theta = \theta_o$ and $\theta = -\theta_o$, satisfying this equation. If $r = 1$, these two values of θ coalesce into a double zero, $\theta = 0$. Different asymptotic expansions are therefore needed for r close to one and r different from one. By an appropriate choice of a the transformation $\sin\theta - r\theta = 3a^2\omega - \omega^3$ can be made analytic for $\theta = \theta_o$ and the expansion of θ in powers of ω remains valid even if r goes through unity. It is then possible to evaluate (19) in terms of Airy integrals and their derivatives. The method is currently being applied to integral representations of the confluent hypergeometric functions are to the evaluation of the Coulomb wave functions.

Another study on the asymptotic evaluation of integrals (BR-9) presents a new method for obtaining an asymptotic representation for integrals of the form

$$\int_0^\infty e^{-px}x^{c-1}f(x)\,dx$$

when p is large. It is shown that if $f(x)$ satisfies certain conditions this representation is also convergent. Numerical calculations seem to show that the first term of the representation gives a close approximation to the value of the integral for a wide range of values of p.

We mention finally two papers on the theory of differential equations which have direct significance for electromagnetic problems. The first of these (BR-5) is essentially an uniqueness theorem for the elliptic partial differential equation and shows that the solutions of $\Delta u + k^2 u = 0$, where k^2 is bounded but not necessarily analytic, vanish identically if they vanish at a point more strongly than any power of the distance to that point. An analogous theorem is proved for the equation $\Delta u = F(x, u)$ provided F satisfies a Lipschitz condition. The second paper (BR-2) discusses the range of applicability of the image method for second-order linear partial differential equations. The image method is used to determine a Green's function for a region of space bounded by planes in terms of the corresponding Green's function for the same operator defined over all space. The equations, regions, and boundary conditions to which the method is applicable are determined

Conclusion

Apologies are no doubt in order for this sketchy and disconnected account of our research. Though, in part, the sketchiness is due to limitations of space and the disconnectedness to the self-will of human minds, these deficiencies are also in part due to the writer. Reference to the papers themselves will, it is to be hoped, clarify much of the vagueness. It is clear, however, that we are greatly indebted to the Air Force Cambridge Research Center and to the Office of Scientific Reasearch of the Air Research and Development Command for generous support of the above work and, more especially, for their farsightedness in attempting to build up a firm theoretical foundation for the multifarious vital applications being made today.

REPORTS AND PUBLICATIONS

*EM-11　W. Sollfrey, "The Variational Solution of Scattering Problems," March, 1949.

*EM-12　A. Leitner, "Notes on Diffraction by a Circular Disk," April, 1949. *Journ. Acous. Soc. Amer.*, vol. 21, p. 331; 1949.

*EM-13　J. B. Keller and H. B. Keller, "Determination of Reflected and Transmitted Fields by Geometrical Optics," May, 1949. *Jour. Opt. Soc. Amer.*, vol. 40, p. 48; 1950.

*EM-14　R. K. Luneberg, "Asymptotic Development of Steady-State Electromagnetic Fields," July, 1949.

*EM-15　R. K. Luneberg, "Asymptotic Evaluation of Diffraction Integrals," October, 1949.

*EM-16　J. B. Keller and H. B. Keller, "A Point Dipole in Spherically Symmetric Media," February, 1950.

*EM-17　B. Friedman, "Numerical Methods for Evaluation of the Integrals for Virtual Height," February, 1950. L. Kraus, *Jour. Geophys. Res.* vol. 58, p. 551; 1953.

**EM-18　J. Shmoys, "Diffraction of Electromagnetic Waves by a Plane Wire Grating," March, 1950. *Jour. Opt. Soc. Amer.*, vol. 41, p. 324; 1951.

EM-19　A. Leitner, "Effect of a Circular Groundplane on Antenna Radiation," April, 1950. *Jour. Appl. Phys.*, vol. 21, p. 1001; 1950.

*EM-20　J. B. Keller and S. Preiser, "Determination of Reflected and Transmitted Fields by Geometrical Optics, Part II," April, 1950.

*EM-21　J. B. Keller and A. Blank, "Diffraction and Reflection of Pulses by Wedges and Corners," June, 1950. *Comm. Pure Appl. Math.*, vol. 4, p. 75; 1951.

EM-22　V. Twersky, "On the Scattered Reflection of Scalar Waves from Absorbent Surfaces," August, 1950. *Jour. Acous. Soc. Amer.*, vol. 23, p. 336; 1951.

*EM-23　S. Karp and W. Sollfrey, "Diffraction by a Dielectric Wedge with Application to Propagation through a Cold Front," October, 1950.

*EM-24　M. Kline, "An Asymptotic Solution of Maxwell's Equations," November, 1950. *Comm. Pure Appl. Math.*, vol. 4, p. 225; 1951.

EM-25　"Separation of Variables and Wiener-Hopf Techniques," December, 1950. *Comm. Pure Appl. Math.*, vol. 3, p. 411; 1950.

*EM-26　V. Twersky, "On the Scattered Reflection of Electromagnetic Waves," January, 1951. *Jour. Appl. Phys.*, vol. 22, p. 825, 1951.

**EM-27　J. Shmoys, "On the Definition of Virtual Height," February, 1951. *Jour. Geophys. Res.*, vol. 57, p. 95; 1952.

*EM-28　B. Friedman, "The Dipole Field in an Inhomogeneous Atmosphere," March, 1951. *Comm. Pure Appl. Math.*, vol. 4, p. 317; 1951.

*EM-29　N. Marcuvitz, "Field Representations in Spherically Stratified Regions," April, 1951. *Comm. Pure Appl. Math.*, vol. 4, p. 263; 1951.

*EM-30　B. Friedman, "Report on a Conference on Dynamics of Ionized Media," May, 1951.

*EM-31　J. Lurye, "Electromagnetic Scattering Matrices of Stratified Anisotropic Media," May, 1951.

EM-32　W. Magnus, "Infinite Matrices Associated with Diffraction by an Aperture," May, 1951. *Quart. Appl. Math.*, vol. 11, p. 77; 1953.

EM-33　H. B. Keller and J. B. Keller, "On Systems of Linear Ordinary Differential Equations," July, 1951.

*EM-34　V. Twersky, "Multiple Scattering of Radiation, Part I; Arbitrary Configuration of Parallel Cylinders, Planar Configuration, Two Cylinders," July, 1951. *Jour. Acous. Soc. Amer.*, vol. 24, p. 42; 1952; *Jour. Appl. Phys.*, vol. 23, p. 407; 1952.

*EM-35　S. N. Karp, "The Natural Charge Distribution and Capacitance of a Finite Conical Shell," September, 1951. Accepted by *Quart. Appl. Math.*

EM-36　J. B. Keller, "Parallel Reflection of Light by Plane Mirrors," October, 1951. *Quart. Appl. Math.*, vol. 11, p. 216; 1953.

*EM-37　"The Theory of Electromagnetic Waves"—A Symposium, November, 1951. *Interscience Publishers, Inc.*, New York, N. Y., December, 1951.

EM-38　A. Russek, "Scattering Matrices for Ionosphere Models," December, 1951.

**EM-39　V. Twersky, "Multiple Scattering of Radiation, Part II, The Grating," December, 1951. *Jour. Appl. Phys.*, vol. 23, p. 1099, 1952—*Jour. Opt. Soc. Amer.*, vol. 42; p. 855; 1952.

*EM-40　W. Magnus, "On the Scattering Effect of a Rough Plane Surface," January, 1952.

EM-41　W. Sollfrey and J. Shmoys, "Diffraction of Electromagnetic Waves by a Plane Wire Grating; II," February, 1952.

*EM-42　R. S. Phillips, "Linear Ordinary Differential Operators of the Second Order," April, 1952.

**EM-43　I. Kay, "Diffraction of an Arbitrary Pulse by a Wedge," April, 1952. *Comm. Pure Appl. Math.*, vol. 4, p. 419; 1953. Appendix by J. B. Keller, *Jour. Appl. Phys.*, vol. 23, p. 1267; 1952.

**EM-44　B. Friedman and J. Russek, "Addition Theorems for Spherical Waves," June, 1952. *Quart. Appl. Math.*, vol. 12, p. 13; 1954.

*EM-45　W. Sollfrey, "Diffraction of Pulses by Conducting Wedges and Cones," July, 1952.

*EM-46　J. Bazer and S. N. Karp, "Propagation of Plane Electromagnetic Waves Past a Shoreline," July, 1952.

*EM-47　B. Friedman, "Techniques of Applied Mathematics—Theory of Distributions," October, 1952.

*EM-48　M. Kline, "An Asymptotic Solution of Linear Second-Order Hyperbolic Differential Equations," December, 1952. *Jour. Rational Mech.*, vol. 3, p. 315; 1954.

*EM-49　J. Russek and V. Twersky, "Graphs of the Function

$$E(N, \delta) = \sum_{n=1}^{N} n^{-1/2} e^{in\delta},$$ February, 1953.

EM-50　C. J. Bouwkamp, "Diffraction Theory—A Critique of Some Recent Developments," April, 1953. *Reports on Progress in Phys.*, vol. 17, p. 35; 1954.

*EM-51　J. Shmoys, "Low-Frequency Propagation in an Exponential Ionospheric Layer," April, 1953.

*EM-52　S. N. Karp, "Diffraction by a Tipped Wedge—with Applications to Blunt Edges," May, 1953.

EM-53　I. Kay, "Diffraction of Pulses by Parabolic Cylinders and Paraboloids of Revolution," June, 1953.

**EM-54　V. Twersky, "Certain Reflection and Transmission Coefficients," June, 1953. *Jour. Appl. Physics.*, vol. 25, p. 859; 1954—*Jour. Appl. Phys.*, vol. 24, p. 659; 1953.

EM-55　I. Kay and J. B. Keller, "Asymptotic Evaluation of the Fields at a Caustic," August, 1953. *Jour. Appl. Phys.*, vol. 25, p. 876; 1954.

EM-56　H. B. Keller, "Ionosphere Propagation of Plane Waves," August, 1953.

EM-57　H. B. Keller, "On the Electromagnetic Field Equations in the Ionosphere," September, 1953.

EM-58　V. Twersky, "Multiple Scattering of Waves by Planar Random Distributions of Parallel Cylinders and Bosses," October, 1953.

EM-59　V. Twersky, "Multiple Scattering of Waves by a Volume Distribution of Parallel Cylinders," October, 1953.

EM-60　B. Friedman, "Techniques of Applied Mathematics—Ordinary Differential Equations and Green's Functions," November, 1953.

EM-61　F. Reiche, "On Diffraction by an Infinite Grating," December, 1953.

EM 62　Müller, "Radiation Patterns and Radiation Fields," March, 1954. *Jour. Rational Mech.*, vol. 4, p. 235, March, 1955.

EM-63　J. Shmoys, "Propagation in Semi-Infinite Waveguides"—Translations of six papers by L. A. Vajnshtejn, January, 1954.

EM-64　F. G. Friedlander, "Diffraction of Pulses by a Circular Cylinder," April, 1954. *Comm. Pure Appl. Math.*, vol. 7, p. 705; 1954.

EM-65　J. Lurye, "The Electromagnetic Field of a Dipole over a Dielectric Slab on a Finitely Conducting Ground Plane," July, 1954.

EM-66　J. Bazer and S. N. Karp, "Potential Flow Through a Conical Pipe with an Application to Diffraction Theory," August, 1954.

EM-67　F. G. Friedlander and J. B. Keller, "Asymptotic Expansions of Solutions of $(\nabla^2 + k^2)u = 0$," September, 1954.

---

* Copies not available for distribution.
** Reprints only available for distribution.

EM–68 J. Meixner, "Diffraction of Electromagnetic Waves by a Slit in a Conducting Plane Between Different Media," October, 1954.

EM–69 N. Marcuvitz, "Field Representations in General Cylindrical Regions. I," November, 1954.

EM–70 W. E. Williams, "Reflection and Refraction of Electromagnetic Waves at Plane Interfaces Between Dielectrics," November, 1954.

EM–71 S. N. Karp, "The Effect of Discontinuities of Dielectric Constant on Fields near Singularities of Conductors," December, 1954.

EM–72 J. Meixner, "The Behavior of Electromagnetic Fields at Edges," December, 1954.

EM–73 N. Chako, "On Integral Relations Involving Products of Spheroidal Functions," January, 1955.

EM–74 I. Kay, "The Inverse Scattering Problem," February, 1955.

EM–75 S. N. Karp and A. Russek, "Diffraction by a Wide Slit," February, 1955.

EM–76 F. G. Friedlander, "Propagation of a Pulse in an Inhomogeneous Medium," March, 1955.

BR–1 ‡ W. Magnus, "Infinite Determinants in the Theory of Mathieu's and Hill's Equations," February, 1953.

BR–2 J. B. Keller, "The Scope of the Image Method," April, 1953. *Comm. Pure Appl. Math.*, vol. 6, p. 505; 1953.

BR–3 W. Magnus, "Algebraic Aspects in the Theory of Systems of Linear Differential Equations," June, 1953. *Comm. Pure Appl. Math.*, vol. 7, p. 649; 1954.

BR–4 (TN–54–23), B. Friedman and L. I. Mishoe, "Eigenfunction Expansions Associated with a Non-Self-Adjoint Differential Equation," January, 1954.

BR–5 (TN–54–52), C. Müller, "The Behavior of the Solutions of $\triangle U = F(x, U)$ in the Neighborhood of a Point," March, 1954. *Comm. Pure Appl. Math.*, vol. 7, p. 505; 1954.

BR–6 (TN–54–141), W. Magnus, "Infinite Determinants Associated with Hill's Equation," June, 1954. Accepted by *Pacific Jour. Math.*

BR–7 (TN–54–195), B. Epstein, "Determination of Coefficients of Capacitance of Regions Bounded by Collinear Slits and of Related Regions," August, 1954.

BR–8 (TN–54–308), W. Magnus, "A Fourier Theorem for Matrices," October, 1954. Accepted by *Proc. Amer. Math. Soc.*

BR–9 (TN–54–342), B. Friedman and J. Franklin, "A Convergent Asymptotic Representation for Integrals," November, 1954. Accepted by *Proc. Camb. Phil. Soc.*

BR–10 M. Hellman, "Lie Algebras Arising from Systems of Linear Differential Equations," March, 1955.

TW–9 B. Friedman, "Amplification of the Traveling-Wave Tube," February, 1949. *Jour. Appl. Phys.*, vol. 22, p. 443; 1951.

TW–10 W. Sollfrey, "Propagation Along a Helical Wire," June, 1949. *Jour. Appl. Phys.*, vol. 22, p. 905; 1951.

TW–11 R. S. Phillips, "The Electromagnetic Field Produced by a Helix," June, 1949. *Quart. Appl. Math.*, vol. 8, p. 229; 1950.

TW–12 B. Friedman and H. Malin, "Excitation of the Traveling-Wave Tube," February, 1950.

TW–13 B. Friedman and H. Malin, "Orthogonality Properties of Modes of the Helical Wave Guide," October, 1950.

TW–14 W. Sollfrey and M. Weitz, "The Gain and Noise Figure of the Helical Traveling-Wave Tube," April, 1951.

TW–15 W. Sollfrey, "Effect of Initial Conditions on Traveling-Wave Tubes," May, 1951.

TW–16 B. Friedman, "Eigenvalues of Compound Matrices," October, 1951.

TW–17 H. Keller and J. Keller, "Eigenvalues of Nearly Circular Waveguides," December, 1951. *Quart. Appl. Math.*, vol. 12, p. 141; 1954.

TW–18 W. Sollfrey, "Magnetic Field Effects in Traveling-Wave Tubes," May, 1952.

TW–19 M. Weitz, "Propagation Constants of Traveling Waves," March, 1953.

TW–20 P. Parzen, "Electromagnetic Wave Propagation in Bounded Electron Beams," May, 1953. *Quart. Appl. Math.*, vol. 12, p. 309; 1954.

170–1 M. Kline, "Relationships between Coaxial and Cylindrical Wave Guide and Cavity Modes"—Also Addendum to 170–1, April, 1947. *Jour. Math. and Phys.*, vol. 27, p. 37; April, 1948.

*170–2 R. S. Phillips, "A Helical Wave Guide," January, 1947.

170–3 R. S. Phillips, "A Helical Wave Guide II," August, 1947. *Amer. Jour. Math.* (Bessel Function Approximations) vol. 72, p. 407; 1950.

170–4 L. Zippin and M. Kline, "Deformation of a Rectangular Guide," December, 1947. *Proc. Amer. Math. Soc.*, vol. 1, p. 543; August, 1950.

170–5 W. Sollfrey, "Equivalent Circuits for Narrow Dielectric Blocks in Wave Guides," January, 1948.

170–6 T. Braverman and J. Lurye, "Bibliography—Wave Guide and Cavity Theory," June, 1948.

170–7 R. S. Phillips and H. Malin, "Investigation of Exceptional Modes $n = \pm 1, \pm 2$," June, 1948. *Amer. Jour. Math.*, (Bessel Function Approximations), vol. 72, p. 407; 1950.

170–8 B. Friedman and H. Malin, "Propagation in Wave Guides Bounded by Anisotropic Plates," July, 1948.

*172–1 J. B. Keller, "Reflection of Electromagnetic Waves," December, 1946.

*172–2 B. Friedman, "Application of Inversion to Wave Propagation," May, 1947.

172–3 R. K. Luneberg, "Propagation of Electromagnetic Waves in Plane Parallel Layers," June, 1947.

172–4 J. Lurye "Propagation of Dipole Radiation through Plane Parallel Layers," November, 1947.

172–5 J. B. Keller, "Reflection and Transmission of Electromagnetic Waves by Thin Curved Shells," February, 1948. *Jour. Appl. Phys.*, vol. 21, p. 896; 1950.

172–6 R. K. Luneberg, "Propagation of Electromagnetic Waves from an Arbitrary Source through Inhomogeneous Stratified Atmospheres," December, 1947.

172–7 H. B. Keller, "Reflection and Transmission of Electromagnetic Waves by a Spherical Shell," January, 1948. *Jour. Appl. Phys.*, vol. 20, p. 393; 1949.

172–8 R. K. Luneberg, "Maxwell's Equations in Spherically Symmetric Media," July, 1948.

172–9 T. Braverman and J. Lurye, "Bibliography—Propagation and Scattering of Electromagnetic Waves," September, 1948.

172–10 J. Greenstadt, "Calculation of Scattering by Thin Shells," November, 1948.

Asymptotic Developments and Scattering Theory in Terms of a Vector Combining the Electric and Magnetic Fields

H. BREMMER†

Summary—The vector combination

$$\overrightarrow{M} = \left(\frac{\mu}{\epsilon}\right)^{1/4}\overrightarrow{H} + j\left(\frac{\epsilon}{\mu}\right)^{1/4}\overrightarrow{E}$$

which was in principle introduced by Bateman and Silberstein in order to shorten Maxwell's equations for homogeneous media, also proves to be useful for the treatment of inhomogeneous media (ϵ and μ not depending on the time). The vector $M\rightarrow$ is to be considered together with its conjugated quantity $M^x\rightarrow$ obtained by replacing the imaginary unit j by $-j$. In a source-free medium the Maxwell equations reduce to

$$\text{curl }\overrightarrow{M} + \frac{j}{c}(\epsilon\mu)^{1/2}\partial\overrightarrow{M}/\partial^f = \tfrac{1}{4}\text{ grad log }\frac{\mu}{\epsilon} \wedge \overrightarrow{M^x}, \quad (1)$$

and to the equation obtained by taking the conjugated complex value.

This relation shows how an interaction between $M\rightarrow$ and $M^x\rightarrow$ is produced only by the inhomogeneity of the medium. The theory of scattering by special volume elements, as well as that of partial reflections against layers with rapidly changing ϵ and μ, can be based on the single relation (1) while fully accounting for the vectorial character of the field. The introduction of $M\rightarrow$ and $M^x\rightarrow$ also enables one to put many results of Luneberg-Kline's theory concerning asymptotic developments in a very simple form. As an example we mention the equation:

$$\text{grad } S \wedge \overrightarrow{m_r} - i(_r\mu)^{1/2}m_r = c\text{ curl }\overrightarrow{m_{r-1}}$$

$$- \frac{c}{\epsilon}\text{ grad log }\frac{\mu}{\epsilon} \wedge \overrightarrow{m_{r-1}{}^x},$$

which fixes all recurrence relations between the consecutive terms of geometric-optical expansions; these expansions are defined by the asymptotic development

$$\overrightarrow{M} = e^{iks}\sum_{r=o}^{\infty}\left(\frac{i}{kc}\right)^r\overrightarrow{m_r}$$

for monochromatic solutions corresponding to some eiconal function S.

MAXWELL'S EQUATION IN TERMS OF BATEMAN'S VECTOR

THE TWO vector equations constituting the Maxwell equations for a homogeneous space were combined into a complex form by Bateman[1] and Silberstein[2]. This enabled the reduction to one single equation, the real and imaginary parts of which yield the ordinary Maxwell equations in the case of real fields. This shortened formulation of Maxwell's equations can be extended to non-homogeneous spaces which we assume for simplicity's sake as sourcefree. In fact, by

defining Bateman's vector in this general case as the combination

$$\overrightarrow{M} = \left(\frac{\mu}{\epsilon}\right)^{1/4}\overrightarrow{H} + j\left(\frac{\epsilon}{\mu}\right)^{1/4}\overrightarrow{E}, \quad \left(j = \sqrt{-1}\right) \quad (1)$$

we find the following representation of Maxwell's equations:

$$\text{curl }\overrightarrow{M} + \frac{j}{c}(\epsilon\mu)^{1/2}\frac{\partial\overrightarrow{M}}{\partial f} = \frac{1}{4}\text{ grad}\cdot\text{log }\frac{\mu}{\epsilon} \wedge \overrightarrow{M^x} \quad (2)$$

(the asterisk denotes the conjugated complex value). This relation, combined with its conjugated complex relation obtained by interchanging j and $-j$ throughout, is fully equivalent to the original Maxwell equations. The relation (2) can also be applied to non-real fields, for instance to monochromatic fields with a time factor $\exp(-i\omega f)$; in this case it is advisable to discriminate between the imaginary unit i in $\partial/\partial f = -i\omega$ and the above imaginary unit $j(i \neq j; i^2 = j^2 = -1)$.

DISCRIMINATION BETWEEN SCATTERING AND INTERNAL REFLECTIONS

The absence of the right-hand side of (2) in homogeneous spaces shows how all coupling between $M\rightarrow$ and $M^x\rightarrow$ can be ascribed to inhomogeneities of the medium. Eq. (2) can be solved by conventional methods in homogeneous spaces owing to the vanishing of its right-hand side, and to the constant value of the refractive index $n = (\epsilon\mu)^{1/2}$ in the left-hand side. In order to facilitate a comparison with the phenomena occurring in homogeneous spaces it is therefore reasonable to split off from n to a constant part n_o which may be some average value of n. This leads to the following representation of (2):

$$\text{curl }\overrightarrow{M} + \frac{j}{c}n_o\frac{\partial\overrightarrow{M}}{\partial f} = -\frac{j}{c}(n - n_o)\frac{\partial\overrightarrow{M}}{\partial f}$$

$$+ \tfrac{1}{4}\text{ grad log }\frac{\mu}{\epsilon} \wedge \overrightarrow{M^x}. \quad (3)$$

If the right-hand side were a known function the equation could be solved completely in the form of some integral extending over the regions for which the right-hand side is nonvanishing. In the actual case, in which the right-hand side depends on the same unknown functions as the left-hand side, this integral solution can also be written out formally; the integral then contains the function yet to be determined. In other words we get an integral equation comprising all information about the field. According to a well-known procedure a first approximation is obtained by substituting in the

[1] H. Bateman, "Electrical and Optical Wave Motion," Cambridge, England, p. 4; 1915.
[2] L. Silberstein, "Elektromagentische Grundgleichungen in bivektorieller Behandlung," *Ann. Phys.* vol. 22, pp. 579–586; 1907. and L. Silberstein, "Nachtrag zur Abhandlung über "Elektromagnetische Grundgleichungen in bivektorieller Behandlung," *Ann. Phys.*, vol. 24, pp. 783–784; 1907.

right-hand side of (3) some solution $M_o\!\rightarrow$ of the homogeneous equation without any right-hand side at all.

The effects caused by the two terms of the right-hand side of (3) can be interpreted as scattering and as internal reflections. In fact, scattering always depends on the local deviations of *n itself* from some average value n_o, whereas internal reflections (i.e. reflections against the infinitely thin layers of constant μ/ϵ composing the actual medium) depend on the spatial *derivative* of μ/ϵ. The scattering part of the first-order approximation is known as the *Born approximation*, whereas the internal reflections lead in special cases to W.K.B. corrections. In the higher-order contributions of the solution of the integral equations scattering and internal reflections may play a combined role.

The Usefulness of the Vector $\overrightarrow{M}$ in Luneberg-Kline's theory

The asymptotic developments of this theory[3] concern the electric and magnetic field; they can be combined for monochromatic fields in a single expansion for the vector $M\!\rightarrow$ according to

$$\overrightarrow{M} = e^{iko} S^{(x,y,z)} \sum_{r=o}^{\infty} \left(\frac{i}{k_o c}\right)^{r} \overrightarrow{m_r}(x, y, z); \qquad (4)$$

the phase function S is some solution of the eiconal equation $|\operatorname{grad} S| = n(x, y, z)$. A formal substitution of (4) into the Maxwell equation (2) (for $\partial/\partial f = -i\omega$) yields the following recurrence relation by assembling the terms of the order $k_o{}^{1-r}$:

$$\operatorname{grad} S \wedge \overrightarrow{m_r} - jnm_r$$
$$= c \operatorname{curl} \overrightarrow{m_{r-1}} - \frac{c}{4} \operatorname{grad} \log \frac{\mu}{\epsilon} \wedge \overrightarrow{m_{r-1}}^x. \qquad (5)$$

This relation is also valid for $r = 0$ if $\overrightarrow{m_{-1}}$ is taken as zero, that is we have:

$$\operatorname{grad} S \wedge \overrightarrow{m_o} = jnm_o. \qquad (6)$$

The vector $\overrightarrow{m_o}$ represents the geometric-optical approximation which is a better approximation according as the wavelength $2\pi/k_o$ becomes smaller. The real and imaginary parts of (6) lead to the well-known triple orthogonality of the geometric-optical approximations $\overrightarrow{e_o}$ and $\overrightarrow{h_o}$ of the electric and magnetic field, and of the vector grad S the direction of which can be identified with the propagation direction of the waves. Further

[3] M. Kline, "An asymptotic solution of Maxwell's equations," *Comm. Appl. Math.*, vol. 4, pp. 225-262; 1951.

properties of $\overrightarrow{m_o}$ can only be found from a further condition, such as that (5) for $r = 1$ (that is a relation depending on both $\overrightarrow{m_o}$ and $\overrightarrow{m_1}$) should be solvable with respect to $\overrightarrow{m_1}$. Luneberg's analysis leads to two so-called "transport equations" which are first-order differential equations for the change of $\overrightarrow{e_o}$ and $\overrightarrow{h_o}$ along a ray-trajectory, that is along a curve which is everywhere perpendicular to the wavefronts $S = \text{constant}$. This author derives two separate transport equations, one referring to $\overrightarrow{e_o}$ and the other to $\overrightarrow{h_o}$2).[4] Here again, the introduction of the bateman vector $\overrightarrow{m}$ simplifies matters; the two equations prove to be equivalent to the real and imaginary parts (with respect to j) of the following single equation for $\overrightarrow{m_o}$:

$$2n \frac{d\overrightarrow{m_o}}{ds} + \left(\Delta S - \frac{dn}{ds}\right)\overrightarrow{m_o}$$
$$+ \frac{2}{n} \operatorname{grad} S \, (\operatorname{grad} n - \overrightarrow{m_o}) = 0; \qquad (7)$$

the derivative d/ds refers to differentiation in the direction of grad S.

The transport equation (7) for $\overrightarrow{m_o}$ was extended by Kline to a general transport equation giving in principle the variation of $\overrightarrow{m_s}$ along a special trajectory in terms of the preceding coefficient $\overrightarrow{m_{s-1}}$ of (4). With the aid of these transport equations the coefficients $\overrightarrow{m_\epsilon}$ can be derived in succession if they are known on some special surface. These rather complicated equations have been expressed in terms of $\overrightarrow{e_r}$ and $\overrightarrow{h_r}$[5] but they very likely become simpler in terms of $\overrightarrow{m_r}$. The latter can be indicated explicitly in the case of a homogeneous space (n constant); it then reads:

$$2n \frac{d\overrightarrow{m_r}}{ds} + \Delta S \cdot \overrightarrow{m_r} = c \cdot \Delta \overrightarrow{m_{r-1}}.$$

We believe that the advantage of the introduction of the M vector in very different problems has been illustrated by the above examples.

[4] R. K. Luneberg, "Asymptotic Development of Steady State Electromagnetic Fields," New York University Math. Res. Group Res. Rep. EM-14 (1949), eq. (7.27).
[5] Kline, *op. cit.*, eqs. (46a) and (46b).

The Theoretical and Numerical Determination of the Radar Cross Section of a Prolate Spheroid

K. M. SIEGEL,† F. V. SCHULTZ,‡ B. H. GERE,§ AND F. B. SLEATOR†

Summary—The exact curve is found for the nose-on radar cross section of a perfectly conducting prolate spheroid whose ratio of major to minor axis is 10:1, for values of π times the major axis divided by the wavelength less than three. The exact acoustical cross section is also found. The mathematical solution is obtained by setting up a series expansion for the scattered wave in terms of two sets of solutions of the vector Helmholtz equation and evaluating the undetermined coefficients in this series by applying the boundary conditions on the surface of the spheroid.

INTRODUCTION

DURING 1949, an effort was begun at the University of Michigan to determine the electromagnetic scattering properties of a perfectly conducting prolate spheroid. The initial problem considered was the determination of the axially symmetric back-scattering cross section of the spheroid. In the first published result,[1] summarized in the Appendix, the method of Hansen was used to determine the algebraic form of the axially symmetric back-scattering cross section in terms of an infinite series of spheroidal wave functions. It was then decided to find the magnitude of the axially symmetric radar cross-section at the first resonance of the prolate spheroid. This problem was set up for computation on a large scale computer, the Mark III electronic digital computer, at the Naval Proving Ground, Dahlgren, Va. The physical reasoning used to predict the region of the first resonance is discussed in Section I; the computational methods and procedures are discussed in the following paper.[2] The preliminary results of these computations were reported at the McGill Symposium[3] and the major analysis was published in December, 1953.[4] In the last year and a half, explanation of the physical basis of these results has been attempted without much success. It is felt that these and the additional results obtained since December, 1953 verify that the physical situation is more complicated than one would conclude from the simple physical picture associated with the interference between an optical scattering contribution and a creeping wave contribution. In order to test the accuracy of the results, new and more precise computations were made

at the first three maxima which occur in the curve of axially symmetric scattering of a sound wave by a 10:1 prolate spheroid. The results still indicate that the theory of a creeping wave plus an optical contribution does not give a complete physical picture for the region near the first few maxima.

The results may suggest that the Rayleigh type contribution still contributes strongly to the amplitude of the cross section beyond the location of the first maximum (wavelength slightly less than 2π times the semi-major axis).

SECTION I

Location of the Resonance Region

In thin wire theory one usually considers the problem of the wavelength large with respect to the diameter of the wire but small with respect to its length. Thus, physically one would expect that the phenomena governing the abscissas of successive maxima for a thin wire would be similar to those pertaining to a thin prolate spheroid. Consequently it was decided that one criterion which would predict approximately the abscissa of the first maximum for the spheroid would be the value given by thin wire theory.

Our experience with the Kirchhoff approximation indicated that it would predict the first maximum at too large a value of $ka(k = 2\pi/\lambda,\ a =$ semi-major axis). This information was plotted with the Rayleigh answer for a thin prolate spheroid ($a/b = 10$, where $b =$ semi-minor axis). These results indicated the necessary spread of points to be computed on the Mark III. The cost, difficulty in computing, and overall size of the problem limited the number of points actually computed to 22.

Results

The results shown in Fig. 1 were obtained from the following formulas. The geometric optics solution is

$$\sigma_{G.O.} = \pi b^4/a^2.$$

The physical optics solution is

$$\sigma_{P.O.} = \frac{\pi b^4}{a^2}\left[1 - 2[\cos ka]\left(\frac{\sin ka}{ka}\right) + \left(\frac{\sin ka}{ka}\right)^2\right].$$

We note that

$$\lim_{k \to \infty} \sigma_{P.O.} = \sigma_{G.O.}$$

We also note that

$$\lim_{k \to 0} \sigma_{P.O.} \sim \frac{\pi b^4}{a^2}[ka]^2 = \sigma_{G.O.} \cdot (ka)^2.$$

The Rayleigh answer is

$$\sigma_{Ray} = \frac{64\pi^3 T^2 k^4}{N^2(4\pi - N)^2},$$

where

$$T = 4\pi a b^2/3,$$

and

$$N = \frac{2\pi a^2}{a^2 - b^2}\left\{1 - \frac{b^2}{2a(a^2 - b^2)^{1/2}}\log\frac{a + (a^2 - b^2)^{1/2}}{a - (a^2 - b^2)^{1/2}}\right\}.$$

For $a = 10b$

$$\sigma_{Ray} \sim \frac{64\pi b^4}{9a^2}(ka)^4 = \frac{64}{9}\cdot\sigma_{G.O.}\cdot(ka)^4.$$

Fig. 1—Electromagnetic back-scattering from a prolate spheroid. $a/b = 10$.

If one considers a 10/1 prolate spheroid as a thin wire one could use the work of Van Vleck, et al.[5] to predict the abscissas of the successive maxima.

The appropriate formula is

$$\frac{\pi}{4}\left\{2\log_e\frac{\lambda}{\pi b} + \tfrac{1}{2}\log_e ka - 1.87\right\}^{-1}$$

$$= \begin{cases}\cot ka & \text{for odd maxima} \\ -\tan ka & \text{for even maxima.}\end{cases}$$

When $a/b = 10$, the left-hand side of this expression reduces to

$$\frac{\pi}{4}\left\{4.12 - 1.5\log_e ka\right\}^{-1}.$$

On the basis of similar analyses for the sphere it seemed probable that four terms in the field expansions would be sufficient to guarantee accuracy of two significant

[5] J. H. Van Vleck, F. Bloch, and M. Hamermesh, "Theory of radar reflection from wires or thin metallic strips", *Jour. Appl. Phys.*, vol 18, p. 274; 1947.

figures in the results for $ka \leqslant 3$. At several values of ka both third- and fourth-order results were obtained in order to give some idea of the magnitude of the errors involved in truncating the infinite determinants. The differences between third and fourth order solutions are small for $ka < 3$.

Difficulty in Interpretation of the Results

The only unexpected characteristic of the results is that the successive maxima *increase*. This seems a little unpalatable from some physical pictures which indicate that the successive maxima and minima are caused by the interference between an optical contribution and a creeping wave contribution. According to this concept, the first maximum occurs (beyond the Rayleigh region) when the optical and creeping wave contributions are in phase and reinforce each other. As the wavelength decreases (keeping the scattering body fixed), a phase difference between the two develops, and the first minimum occurs where this reaches 180 degrees. The difficulty arises at the second maximum. As the wavelength continues to decrease, the creeping wave has to travel a larger distance in wavelengths. The creeping wave contribution is exponentially damped by a factor of kl_1 raised to a positive power (l_1 is the distance the wave travels around the body, in wavelengths). Thus it would be expected that the creeping wave contribution to the successive in-phase maxima would decrease.

If the physical optics approximation is in any measure indicative of what happens to the optics contribution in the Franz-Imai theory, it follows that the optics contribution should be slowly varying compared with the creeping wave contribution, and in fact should approach the geometric optics result,

$$\frac{\pi b^4}{a^2}, \quad \text{as} \quad ka \to \infty.$$

Thus as λ decreases the successive in-phase maxima of the total result would be expected to decrease. The results obtained show that this is not the case!

Since the error in the results computed by the Mark III is an increasing function of ka, the significance of the results might be questioned when the third maximum is reached. To check this phenomenon much more detailed computations were made for the acoustic scattering problem.

Exact Acoustics Result

The exact acoustics result is given in Fig. 2. The Rayleigh acoustic answer is

$$\sigma_{a.Ray} = \frac{4\pi^3 T^2}{\lambda^4}\left(\frac{2 - L}{1 - L}\right)^2,$$

where

$$L = \frac{b^2}{a^2 - b^2}\left[\frac{a}{2(a^2 - b^2)^{1/2}}\log\frac{a + (a^2 - b^2)^{1/2}}{a - (a^2 - b^2)^{1/2}} - 1\right].$$

For $a/b = 10$

$$\sigma_{a.Ray} \sim \frac{16}{9}\frac{\pi b^4}{a^2}(ka)^4,$$

Fig. 2—Acoustical back-scattering from a prolate spheroid. $a/b = 10$.

eka terms ($e = 2.718$) to insure two-significant-figure accuracy.

Discussion of the Acoustic Results

We observe that the ordinate of the initial maximum in the acoustic result is only slightly greater than one. If the successive maxima always decrease, as they do in the sphere answer, then since asymptotically they must approach one, it is clear that the changes in magnitude of the maximum will only occur in the third significant figure and thus one may predict immediately to two-significant-figure accuracy, that the ordinate of any maximum beyond the first is equal to 1.0. If one of these maxima does become less than one then again the creeping wave theory would be questioned. One may also predict that even for moderately large ka the scattering cross-section will probably not oscillate symmetrically about the geometric optics result.

Discussion of the Electromagnetic Results

Consider the behavior of the sphere answer as the wavelength decreases with the radius of the body fixed. In the region of very large wavelength the exact answer is in agreement with the Rayleigh dipole answer. Presently, as the wavelength approaches $2\pi a_1$ (a_1 is the radius of the sphere), a deviation from the Rayleigh dipole result appears. This indicates that the true dipole coefficient differs from the Rayleigh dipole coefficient and that the contributions from higher-order poles are now probably significant. The curve passes smoothly through a maximum and starts descending smoothly toward a minimum, implying that the magnitude of the true dipole coefficient changes more rapidly with λ than does the phase of the total contribution from the higher-order poles. If this were not the case, secondary oscillations would appear superimposed on the primary ones.

Suppose, on the other hand, that the difference between the true dipole coefficient and the Rayleigh type dipole coefficient changes slowly with λ as compared to the rate at which the phase changes, and suppose that the magnitudes and phases of the contributions from higher-order poles vary slowly with λ. Then one would expect that the successive maxima of the total field would increase as λ decreases until the dipole contribution becomes negligible. This provides a possible explanation of the increase observed in the first few maxima of the prolate spheroid's nose-on cross section. Note that this would explain both the increase in the maxima and the increased decay from the Rayleigh answer.

Since for the prolate spheroid the first maximum occurs at a wavelength slightly less than $2\pi a$, and since the successive maxima do not immediately decrease, one may reason that the contribution of the dipole (or dipoles) does not decay rapidly as a function of λ.

On the other hand there is some doubt whether there are enough terms present to guarantee that the second and third maxima do really rise above the first maximum. If the analysis of the truncation error in the scalar

which is very nearly 1/4 of the Rayleigh electromagnetic answer. Fig. 2 also contains points previously computed by Spence and Granger[6] to exhibit a comparison of results.

Table I contains, for 7 values of ka, the axially symmetric differential back-scattering cross section divided by the geometric optics cross section as a function of the truncation parameter l. In the computations much more emphasis was placed on the maxima than on the minima.

TABLE I: $\dfrac{\sigma}{\pi b^4 / a^2}$

l	$ka=1.4$	$ka=1.6$	$ka=2.8$	$ka=3.0$	$ka=3.015$	$ka=4.3$	$ka=4.6$
0	0.4921	0.5778	0.2498	0.1700	0.172	0.008714	0.003588
1	1.516	1.530	0.08858	0.3166	0.337	0.7366	0.49232
2	1.0835	0.8821	1.585	2.177	2.211	0.001115	0.32985
3	1.1026	0.9184	0.7577	0.8936	0.885	1.8672	3.1898
4	1.1022	0.9184	0.8201	1.005	1.002	0.53655	0.72876
5	1.1022	0.9185	0.8174	1.001	0.996	0.66209	1.0106
6	1.1022	0.9185	0.8174	1.001	0.996	0.65336	0.98648
7	1.1022	0.9185	0.8174	1.001	0.996	0.65373	0.98768
8	1.1022	0.9185	0.8174	1.001	0.996	0.65372	0.9875

This function exhibits a phenomenon similar to that reported by Debye[7] for the sphere, namely that in this range of ka one need only keep $ka + 2$ terms. In fact for all ka, Ritt[8] has shown that one never needs more than

[6] R. D. Spence and S. Granger, "Scattering of sound from a prolate spheroid," *Jour. Acoust. Soc. Amer.*, vol. 23, p. 701; 1951.
[7] P. Debye, *Ann. der Physik*, vol. 30, p. 57; 1909.
[8] R. K. Ritt, private communication, 1953.

Fig. 3—Prolate spheroidal coordinates.

problem is indicative of the vector result, one is led to believe that there is *at least* no measurable decrease as one goes from the first maximum to the second.

Conclusion

This paper presents the exact nose-on cross section for a 10:1 prolate spheroid. For the electromagnetic problem, it is believed the results are correct to two significant figures up to $ka = 2$.

The exact nose-on acoustic cross section is obtained and it is believed the results are correct to two significant figures up to $ka = 4.6$. If one assumes that the successive maxima always decrease then one can predict in the acoustical case that all maxima past the first have as ordinate the value 1.0 to two significant figures.

APPENDIX

DETERMINATION OF MATHEMATICAL SOLUTION

I. Method of Solution

The theoretical solution to the problem was obtained in the following manner.

A plane, monochromatic electromagnetic wave with time dependence of the form e^{+jwt} (where $j = \sqrt{-1}$) is assumed to strike nose-on a perfectly conducting prolate spheroid embedded in free space. A prolate spheroidal coordinate system is employed, which is related to a rectangular system by the equations

$$x = F\sqrt{(\xi^2 - 1)(1 - \eta^2)} \cdot \cos \phi \qquad (1)$$

$$y = F\sqrt{(\xi^2 - 1)(1 - \eta^2)} \cdot \sin \phi \qquad (2)$$

$$z = F\xi\eta, \qquad (3)$$

where F is the semi-focal distance of the spheroidal system, (Fig. 3).

The expression for the electric vector of the scattered wave must meet the following conditions:

1. It must satisfy the vector wave equation (or the vector Helmholtz equation for a monochromatic wave, as we have here).
2. Its divergence must be zero.
3. At very large distances from the prolate spheroid the scattered wave must take on the behavior of a spherical diverging wave with the center of the spheroid as its center.
4. The resultant of the incident and scattered waves must satisfy the proper boundary conditions over the surface of the spheroid.

In order to obtain an expression for the electric vector of the scattered wave which meets the listed conditions, the following procedures are used:

a. Solutions of the vector Helmholtz equation in prolate spheroidal coordinates are formed from solutions of the scalar Helmholtz equation by means of the method described by Stratton.[9] Solutions are formed which have zero divergence.

b. The incident wave is expressed in terms of appropriate solutions of the vector Helmholtz equation in prolate spheroidal coordinates.

c. The electric vector of the scattered wave is expressed as an infinite series, with undetermined coefficients, of appropriate solutions of the vector Helmholtz equation which satisfy condition 3 above.

d. The undetermined coefficients in the series for the electric vector of the scattered wave are evaluated with the aid of the boundary condition on the surface of the scattering prolate spheroid.

II. *Solutions of the Vector Helmholtz Equation*

As mentioned previously, solutions of the vector Helmholtz equation are to be formed from solutions of the scalar Helmholtz equation by means of the procedure set forth by Stratton.

Solutions of the scalar Helmholtz equation,

$$\nabla^2 \psi + k^2 \psi = 0, \tag{4}$$

are used as presented by Stratton, Morse, Chu, and Hutner.[10] When (4) is solved by the method of separation of variables, three second-order linear ordinary differential equations result: one in η, one in ξ, and one in ϕ. The separation constants are m and n.

Solutions of the equation in ϕ are $e^{jm\phi}$, $\sin m\phi$, and $\cos m\phi$. Obviously m must be an integer in the present work.

Because of the range of $\eta (-1 \leq \eta \leq +1)$, only the first solution of the equation in η may be used, since only this solution is regular throughout the range of η. The solution given by Stratton, Morse, Chu, and Hutner is

$$S_{mn}^{(1)}(\eta) = \sum_{k=0,1}^{\infty\prime} d_k^{mn} P_{k+m}^m(\eta). \tag{5}$$

The numerical coefficients d_k^{mn} are tabulated by Stratton, Morse, Chu, and Hutner for values of $c = (2\pi/\lambda)F = kF$ from 0 to 5 and for values of $m + n$ from 0 to 3. The prime on the summation symbol indicates that the summation is to be over even values of k if n is even, and over odd values of k if n is odd. The functions $P_{k+m}^m(\eta)$ are associated Legendre functions.

The variable ξ has the range $1 \leq \xi < \infty$. Stratton, Morse, Chu, and Hutner give the following solution which is regular for all finite values of ξ:

[9] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York.; N. Y. 1941.

[10] J. A. Stratton, P. M. Morse, L. J. Chu, and R. A. Hutner, "Elliptic Cylinder and Spheroidal Wave Functions," John Wiley & Sons, Inc., New York, N. Y.; 1941.

$$R_{mn}^{(1)}(\xi) = \frac{(\xi^2-1)^{m/2}}{\xi^m \sum\limits_{k=0,1}^{\infty\prime} d_k^{mn} \frac{(k+2m)!}{k!}} \sum_{k=0,1}^{\infty\prime} j^{n-k} d_k^{mn} \frac{(k+2m)!}{k!} j_{m+k}(c\xi), \tag{6}$$

where $j_{m+k}(c\xi)$ is the spherical Bessel function of the first kind. It is to be observed that the coefficients d_k^{mn} used in (6) are the same as those used in (5). A second solution which has logarithmic singularities at $\xi = \pm 1$ is also given,

$$R_{mn}^{(2)}(\xi) = \frac{(\xi^2-1)^{m/2}}{\xi^m \sum\limits_{k=0,1}^{\infty\prime} d_k^{mn} \frac{(k+2m)!}{k!}} \sum_{k=0,1}^{\infty\prime} j^{n-k} d_k^{mn} \frac{(k+2m)!}{k!} n_{m+k}(c\xi), \tag{7}$$

where $n_{m+k}(c\xi)$ is the spherical Bessel function of the second kind.

When the above expression for $R_{mn}^{(2)}(\xi)$ converges very slowly ($\xi \approx 1$ or m large), the following equations may be used:

n even:

$$R_{mn}^{(2)}(\xi) = \left[\frac{2c^{m-1}\Gamma\left(\dfrac{n+2m+1}{2}\right)}{\Gamma\left(\dfrac{n+2}{2}\right)\Gamma(m-\frac{1}{2}) d_{-2m}^{mn} \sum\limits_{k=0}^{\infty\prime} d_k^{mn} \dfrac{(k+2m)!}{k!}} \right] \sum_{k=-\infty}^{+\infty\prime} d_k^{mn} Q_{m+k}^m(\xi), \tag{8}$$

n odd:

$$R_{mn}^{(2)}(\xi) = \left[\frac{-8c^{m-2}\Gamma\left(\dfrac{n+2m+2}{2}\right)}{\Gamma\left(\dfrac{n+1}{2}\right)\Gamma\left(m-\dfrac{3}{2}\right) d_{1-2m}^{mn} \sum\limits_{k=1}^{\infty\prime} d_k^{mn} \dfrac{(k+2m)!}{k!}} \right] \sum_{k=-\infty}^{+\infty\prime} d_k^{mn} Q_{m+k}^m(\xi); \tag{9}$$

The functions $Q_{m+k}^m(\xi)$ are Legendre functions of the second kind.

Two other useful solutions of the differential equation in ξ can be formed as follows:

$$R_{mn}^{(3)}(\xi) = R_{mn}^{(1)}(\xi) + jR_{mn}^{(2)}(\xi), \tag{10}$$

$$R_{mn}^{(4)}(\xi) = R_{mn}^{(1)}(\xi) - jR_{mn}^{(2)}(\xi). \tag{11}$$

These are noteworthy for having the following asymptotic behavior for large $c\xi$:

$$R_{mn}^{(3)}(\xi) \approx \frac{1}{c\xi} e^{j\left(c\xi - \frac{n+m+1}{2}\pi\right)}, \tag{12}$$

$$R_{mn}^{(4)}(\xi) \approx \frac{1}{c\xi} e^{-j\left(c\xi - \frac{n+m+1}{2}\pi\right)}. \tag{13}$$

The solutions of (4) which are used in the present paper are

$$\psi_{mn}^{(h)} = S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi) \cos m\phi. \qquad (14)$$

The superscript "h" may take the values 1, 2, 3, or 4.

From these solutions of the scalar Helmholtz equation the following solutions of the vector Helmholtz equation may be formed (a being an arbitrary constant vector):

$$\underline{L}_{mn}^{(h)} = \nabla \psi_{mn}^{(h)}, \qquad (15)$$

$$\underline{M}_{mn}^{(h)} = \underline{L}_{mn}^{(h)} \times \underline{a} = \nabla \psi_{mn}^{(h)} \times \underline{a}, \qquad (16)$$

$$\underline{N}_{mn}^{(h)} = \frac{1}{k} \nabla \times \underline{M}_{mn}^{(h)}. \qquad (17)$$

Series in terms of $\underline{L}_{mn}$, $\underline{M}_{mn}$, and $\underline{N}_{mn}$ may be used to express any ordinary solution of the vector wave equation. Only $\underline{M}_{mn}$ and $\underline{N}_{mn}$ are of use here since the propagation of electromagnetic waves in free space is under consideration. This physical condition requires that

$$\nabla \cdot \underline{E} = 0, \qquad (18)$$

$$\nabla \cdot \underline{H} = 0; \qquad (19)$$

and from the defining equations given above it is clear that

$$\nabla \cdot \underline{L}_{mn}^{(h)} \neq 0, \qquad (20)$$

$$\nabla \cdot \underline{M}_{mn}^{(h)} = \nabla \cdot \underline{N}_{mn}^{(h)} = 0. \qquad (21)$$

The vectors $\underline{E}$ and $\underline{H}$ can be expressed in infinite series of the functions $\underline{M}_{mn}^{(h)}$ and $\underline{N}_{mn}^{(h)}$.

III. Expression for Incident Wave

Let it be assumed that the incident plane wave is moving along the z-axis in the negative z-direction, that the electric vector has a magnitude E_o and points in the positive y-direction, and that the magnetic vector has a magnitude H_o and points in the positive x-direction. This is the situation for a plane electromagnetic wave striking the spheroid nose-on.

With the help of an expansion developed by Morse[11] the incident electric and magnetic field vectors may be expressed as follows:

$$^{I}\underline{E} = -j \frac{^{I}E_0}{k} \sum_{n=0}^{\infty} A_{on} {}^{x}\underline{M}_{on}^{(1)}, \qquad (22)$$

$$^{I}\underline{H} = -j \frac{^{I}H_0}{k} \sum_{n=0}^{\infty} A_{on} {}^{y}\underline{M}_{on}^{(1)}, \qquad (23)$$

where

$$A_{on} = \frac{2j^n}{N_{on}} \sum_{k=0,1}^{\infty'} d_k^{on}, \qquad (24)$$

$$N_{on} = \sum_{k=0,1}^{\infty'} \frac{2}{2k+1} (d_k^{on})^2. \qquad (25)$$

The vector expressions $^{x}\underline{M}_{m}^{(1)}$, $^{y}\underline{M}_{m}^{(1)}$, $^{z}\underline{M}_{m}^{(1)}$ are found as follows. By definition,

$$^{x}\underline{M}_{m}^{(h)} = \nabla \psi_{mn}^{(h)} \times \underline{i}_x, \qquad (26)$$

$$^{y}\underline{M}_{m}^{(h)} = \nabla \psi_{mn}^{(h)} \times \underline{i}_y, \qquad (27)$$

$$^{z}\underline{M}_{m}^{(h)} = \nabla \psi_{mn}^{(h)} \times \underline{i}_z, \qquad (28)$$

where $\underline{i}_x$, $\underline{i}_y$, and $\underline{i}_z$ are the unit vectors in the x-, y-, and z-directions, respectively. For $h = 1, 2, 3,$ or 4, the function $\psi_{mn}^{(h)}$ is expressed in prolate spheroidal coordinates and its gradient is found.[12] The unit vector $\underline{i}_x$, $\underline{i}_y$, or $\underline{i}_z$ is also expressed in prolate spheroidal coordinates. The results are:

$$^{x}\underline{M}_{mn}^{(h)}$$
$$= \underline{i}_\eta F^{-1}\{ m\xi(\xi^2 - 1)^{-1/2}(\xi^2 - \eta^2)^{-1/2} S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi)$$
$$\cos \phi \sin m\phi$$
$$- (\xi^2 - 1)^{1/2}(\xi^2 - \eta^2)^{-1/2} S_{mn}^{(1)}(\eta) \left[\frac{d}{d\xi} R_{mn}^{(h)}(\xi) \right]$$
$$\sin \phi \cos m\phi \}$$
$$+ \underline{i}_\xi F^{-1}\{ (1 - \eta^2)^{1/2}(\xi^2 - \eta^2)^{-1/2} \left[\frac{d}{d\eta} S_{mn}^{(1)}(\eta) \right] R_{mn}^{(h)}(\xi)$$
$$\sin \phi \cos m\phi$$
$$+ m\eta(1 - \eta^2)^{-1/2}(\xi^2 - \eta^2)^{-1/2} S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi)$$
$$\cos \phi \sin m\phi \}$$
$$+ \underline{i}_\phi F^{-1}\{ \eta(\xi^2 - 1)(\xi^2 - \eta^2)^{-1} S_{mn}^{(1)}(\eta) \left[\frac{d}{d\xi} R_{mn}^{(h)}(\xi) \right]$$
$$\cos \phi \cos m\phi$$
$$+ \xi(1 - \eta^2)(\xi^2 - \eta^2)^{-1} \left[\frac{d}{d\eta} S_{mn}^{(1)}(\eta) \right] R_{mn}^{(h)}(\xi)$$
$$\cos \phi \cos m\phi \}. \qquad (29)$$

$$^{z}\underline{M}_{mn}^{(h)}$$
$$= \underline{i}_\eta F^{-1}\{ m\eta(1 - \eta^2)^{-1/2}(\xi^2 - \eta^2)^{-1/2} S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi)$$
$$\sin m\phi \}$$
$$+ \underline{i}_\xi F^{-1}\{ -m\xi(\xi^2 - 1)^{-1/2}(\xi^2 - \eta^2)^{-1/2} S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi)$$
$$\sin m\phi \}$$
$$+ \underline{i}_\phi F^{-1}\{ -\xi(\xi^2 - 1)^{1/2}(1 - \eta^2)^{1/2}(\xi^2 - \eta^2)^{-1} S_{mn}^{(1)}(\eta)$$
$$\left[\frac{d}{d\xi} R_{mn}^{(h)}(\xi) \right] \cos m\phi$$
$$+ \eta(\xi^2 - 1)^{1/2}(1 - \eta^2)^{1/2}(\xi^2 - \eta^2)^{-1}$$
$$\left[\frac{d}{d\eta} S_{mn}^{(1)}(\eta) \right] R_{mn}^{(h)}(\xi) \cos m\phi \}. \qquad (30)$$

The expression for $^{y}\underline{M}_{mn}^{(h)}$ is not listed since it is not used in the present work.

IV. Expression for Scattered Wave

The M-vectors used in Section III to express the incident waves satisfy the vector Helmholtz equation and have zero divergence, as is pointed out in Section II. Therefore, if the scattered wave is expressed as a linear combination of these vectors with constant coefficients, the resulting expression will satisfy conditions 1 and 2 of Section I. Actually an infinite series of M-vectors, with undetermined coefficients, will be used for expressing the scattered wave. For the representation of the M-vectors, the following solution of the scalar Helmholtz equation (in prolate spheroidal coordinates) is

---

[11] P. M. Morse, "Addition Formulae for Spheroidal Functions," *Proc. Nat. Acad. Sci.*, vol. 21, pp. 56–62; 1935.

[12] Stratton, *Op. cit.*, p. 49, eq. (78).

used:

$$\psi_{mn}^{(h)} = S_{mn}^{(1)}(\eta) R_{mn}^{(h)}(\xi) \cos m\phi. \tag{31}$$

As was pointed out in Section II, four different forms of $R_{mn}^{(h)}(\xi)$ are available and their pertinent properties are given there. Since the incident plane wave exists throughout finite space, in the expression for it given by (22) it is necessary to use the function $R_{mn}^{(1)}(\xi)$, which is regular for all finite values of ξ. The scattered wave exists only outside the perfectly conducting spheroid, where $1 < \xi < \infty$, assuming the minor axis is different from zero. Each of the four forms of $R_{mn}^{(h)}(\xi)$ is regular in this range, but only $R_{mn}^{(4)}$ has the requisite asymptotic behavior for very large values of $c\xi$, as given by (13). Consequently the function $\psi_{mn}^{(4)}$ is used to represent the scattered wave. The final expression for the scattered wave will then satisfy condition 3 of Appendix Section I.

Condition 4 of Appendix Section I will be satisfied since it is used to calculate the coefficients in the series for the scattered wave. The prolate spheroid is assumed to be perfectly conducting, so that this boundary condition is expressed by

$$[\underline{n} \times {}^T\underline{E}]_{\xi=\xi_0} = [\underline{i}_\xi \times ({}^I\underline{E} + {}^S\underline{E})]_{\xi=\xi_0} = 0, \tag{32}$$

where ${}^I\underline{E}$ is the electric vector of the total electromagnetic field and $\underline{n}$ is the unit vector normal to the surface of the prolate spheroid. The number ξ_0 is the value of ξ on the surface of the spheroid. The boundary condition implies that

$$[{}^I E_\eta + {}^S E_\eta]_{\xi=\xi_0} = 0, \tag{33}$$

$$[{}^I E_\phi + {}^S E_\phi]_{\xi=\xi_0} = 0, \tag{34}$$

where ${}^I E_\eta$ is the η-component of ${}^I\underline{E}$ and ${}^I E_\phi$ is the ϕ-component, and similarly for ${}^S\underline{E}$.

Since there are two boundary equations to be satisfied, it is to be expected that it will be necessary to use for ${}^S\underline{E}$ the sum of two series of vector functions, each with undetermined coefficients. At least eight sets of such vector functions are available, the individual terms of each set satisfying the vector Helmholtz equation and having zero divergence. These sets are

$$\begin{aligned} &{}^x\underline{M}_{mn}^{(h)},\ {}^y\underline{M}_{mn}^{(h)},\ {}^z\underline{M}_{mn}^{(h)},\ {}^r\underline{M}_{mn}^{(h)},\ {}^x\underline{N}_{mn}^{(h)}, \\ &\qquad {}^y\underline{N}_{mn}^{(h)},\ {}^z\underline{N}_{mn}^{(h)},\ {}^r\underline{N}_{mn}^{(h)}. \end{aligned} \tag{35}$$

Two of these sets of vector functions have not been previously discussed. They are

$$ {}^r\underline{M}_{mn}^{(h)} = \nabla\psi_{mn}^{(h)} \times \underline{i}_r, \tag{36}$$

and

$$ {}^r\underline{N}_{mn}^{(h)} = \frac{1}{k}\nabla \times {}^r\underline{M}_{mn}^{(h)}, \tag{37}$$

where $\underline{i}_r$ is the unit radius vector. Of course $\underline{i}_r$ is not a *constant* vector but Stratton[9] shows that nevertheless ${}^r\underline{M}$ and ${}^r\underline{N}$ satisfy the vector Helmholtz equation. In choosing two of the eight listed sets of vector functions to represent the scattered wave, one chooses functions whose three components vary with ϕ in the same way as do the corresponding components of the incident wave. For the E-vector of the scattered wave the vector func-

tions which have the proper variation with ϕ are

$$ {}^x\underline{M}_{on}^{(h)},\ {}^z\underline{M}_{1n}^{(h)},\ {}^z\underline{N}_{1n}^{(h)},\ {}^r\underline{M}_{1n}^{(h)},\ {}^r\underline{N}_{1n}^{(h)}. \tag{38}$$

Of these, ${}^x\underline{M}_{on}^{(4)}$ and ${}^z\underline{M}_{1n}^{(4)}$ were chosen for reasons of simplicity. The resulting expression for the scattered wave is

$$ {}^S\underline{E} = \sum_{n=0}^{\infty} [\alpha_n\, {}^x\underline{M}_{on}^{(4)} + \beta_n\, {}^z\underline{M}_{1n}^{(4)}] \tag{39}$$

where α_n and β_n are undetermined coefficients. Expressions for ${}^x M_{on}^{(4)}$ and ${}^z M_{1n}^{(4)}$ are given by (29) and (30), respectively.

It can be shown that, for large $c\xi$,

$$\left[\frac{d}{d\xi} R_{mn}^{(4)}(\xi)\right] \approx \frac{1}{\xi}\, e^{-j(c\xi - n + m/2\,\pi)}. \tag{40}$$

From (13) and (40) results the asymptotic expression

$$\begin{aligned} {}^x\underline{M}_{on}^{(4)} &\approx \big[\, -\underline{i}_\eta S_{on}^{(1)}(\eta) \sin\phi \\ &\quad + \underline{i}_\phi \eta\, S_{on}^{(1)}(\eta) \cos\phi\big]\, \frac{1}{F\xi}\, e^{-j\left(c\xi - \frac{[n+m]}{2}\pi\right)}, \end{aligned} \tag{41}$$

for large $c\xi$.

Now

$$ c\xi = \frac{2\pi}{\lambda} F\xi \tag{42}$$

and

$$\lim_{c\xi \to \infty} (F\xi) = r, \tag{43}$$

so that (41) becomes

$$\begin{aligned} {}^x M_{on}^{(4)} &\approx \big[-\underline{i}_\eta S_{on}^{(1)}(\eta) \sin\phi \\ &\quad + \underline{i}_\phi \eta S_{on}^{(1)}(\eta) \cos\phi\big]\frac{1}{r} e^{-j(2\pi r/\lambda - n\pi/2)}. \end{aligned} \tag{44}$$

In a similar way it may be shown that

$$\begin{aligned} {}^z\underline{M}_{1n}^{(4)} &\approx \big[-\underline{i}_\phi(1 - \eta^2)^{1/2} \\ &\quad S_{1n}^{(1)}(\eta)\ \cos\phi\big]\frac{1}{r}\, e^{-j[2\pi r/\lambda - (n+1/2)\pi]}. \end{aligned} \tag{45}$$

Eqs. (44) and (45) show that each term of (39) for ${}^S\underline{E}$ has the correct behavior at large distances from the prolate spheroid.

The expression for ${}^S\underline{E}$, (39), therefore satisfies the first three of the four necessary conditions. The fourth requirement, that of satisfying the boundary conditions over the surface of the spheroid, is met when these conditions are used, in the next section, to determine the coefficients α_n and β_n of (39).

V. Satisfying Boundary Conditions

The boundary conditions are expressed by (33) and (34). If one substitutes the expressions for ${}^I\underline{E}$ and ${}^S\underline{E}$ given by (22) and (39), respectively, into (33) there results

$$\begin{aligned} &j\frac{{}^I E_o}{k} \sum_{n=0}^{\infty} A_n(\xi_0^2 - 1)^{1/2} S_{on}^{(1)}(\eta)\left[\frac{d}{d\xi} R_{on}^{(1)}(\xi)\right]_{\xi=\xi_0} \\ &= \sum_{n=0}^{\infty} \alpha_n(\xi_0^2 - 1)^{1/2} S_{on}^{(1)}(\eta)\left[\frac{d}{d\xi} R_{on}^{(4)}(\xi)\right]_{\xi=\xi_0} \\ &\quad - \sum_{n=0}^{\infty} \beta_n\eta(1 - \eta^2)^{-1/2} S_{1n}^{(1)}(\eta) R_{1n}^{(4)}(\xi_0). \end{aligned} \tag{46}$$

Similarly, by using (34), one obtains

$$j\frac{^IE_o}{k}\sum_{n=0}^{\infty}A_n\Big\{\eta(\xi_o^2-1)S_{on}^{(1)}(\eta)\Big[\frac{d}{d\xi}R_{on}^{(1)}(\xi)\Big]_{\xi=\xi_o}$$

$$+\xi_o(1-\eta^2)\Big[\frac{d}{d\eta}S_{on}^{(1)}(\eta)\Big]R_{on}^{(1)}(\xi_o)\Big\}$$

$$=\sum_{n=0}^{\infty}\alpha_n\Big\{\eta(\xi_o^2-1)S_{on}^{(1)}(\eta)\Big[\frac{d}{d\xi}R_{on}^{(4)}(\xi)\Big]_{\xi=\xi_o}$$

$$+\xi_o(1-\eta^2)\Big[\frac{d}{d\eta}S_{on}^{(1)}(\eta)\Big]R_{on}^{(4)}(\xi_o)\Big\}$$

$$-\sum_{n=0}^{\infty}\beta_n\Big\{\xi_o(\xi_o^2-1)^{1/2}(1-\eta^2)^{1/2}S_{1n}^{(1)}(\eta)\Big[\frac{d}{d\xi}R_{1n}^{(4)}(\xi)\Big]_{\xi=\xi_o}$$

$$-\eta(\xi_o^2-1)^{1/2}(1-\eta^2)^{1/2}\Big[\frac{d}{d\eta}S_{1n}^{(1)}(\eta)\Big]R_{1n}^{(4)}(\xi_o)\Big\}.\quad(47)$$

These equations must hold for all allowed values of η and may be used to determine α_n and β_n. Eqs. (46) and (47) are multiplied by $S_{ON}^{(1)}(\eta)$ (where N is any nonnegative integer) and then integrated from $\eta=-1$ to $\eta=+1$. The following results are obtained:

$$(^IE_oa)\Big(\frac{j}{c\xi_o}\Big)\sum_{n=0}^{\infty}A_n(\xi_o^2-1)^{1/2}\Big[\frac{d}{d\xi}R_{on}^{(1)}(\xi)\Big]_{\xi=\xi_o}$$

$$\int_{-1}^{+1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta=\sum_{n=0}^{\infty}\alpha_n(\xi_o^2-1)^{1/2}$$

$$\Big[\frac{d}{d\xi}R_{on}^{(4)}(\xi)\Big]_{\xi=\xi_o}\int_{-1}^{+1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta$$

$$-\sum_{n=0}^{\infty}\beta_nR_{1n}^{(4)}(\xi_o)\int_{-1}^{+1}\frac{\eta}{\sqrt{1-\eta^2}}S_{1n}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta,\quad(48)$$

$$(^IE_oa)\Big(\frac{j}{c\xi_o}\Big)\sum_{n=0}^{\infty}A_n\Big\{(\xi_o^2-1)\Big[\frac{d}{d\xi}R_{on}^{(1)}(\xi)\Big]_{\xi=\xi_o}$$

$$\int_{-1}^{+1}\eta S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta+\xi_oR_{on}^{(1)}(\xi_o)$$

$$\int_{-1}^{+1}(1-\eta^2)\Big[\frac{d}{d\eta}S_{on}^{(1)}(\eta)\Big]S_{oN}^{(1)}(\eta)\,d\eta\Big\}$$

$$=\sum_{n=0}^{\infty}\alpha_n\Big\{(\xi_o^2-1)\Big[\frac{d}{d\xi}R_{on}^{(4)}(\xi)\Big]_{\xi=\xi_o}$$

$$\int_{-1}^{+1}\eta\,S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)\,d\eta$$

$$+\xi_oR_{on}^4(\xi_o)\int_{-1}^{+1}(1-\eta^2)\Big[\frac{d}{d\eta}S_{on}^{(1)}(\eta)\Big]S_{oN}^{(1)}(\eta)\,d\eta\Big\}$$

$$-\sum_{n=0}^{\infty}\beta_n\Big\{\xi_o(\xi_o^2-1)^{1/2}\Big[\frac{d}{d\xi}R_{1n}^{(4)}(\xi)\Big]_{\xi=\xi_o}$$

$$\int_{-1}^{+1}\sqrt{1-\eta^2}\,S_{on}^{(1)}(\eta)\,S_{oN}^{(1)}(\eta)\,d\eta-(\xi_o^2-1)^{1/2}$$

$$R_{1n}^{(4)}(\xi_o)\int_{-1}^{+1}\eta\sqrt{1-\eta^2}\Big[\frac{d}{d\eta}S_{1n}^{(1)}(\eta)\Big]S_{oN}^{(1)}(\eta)\,d\eta\Big\}.$$

$$(49)$$

Stratton, Morse, Chu, and Hutner[10] show that the functions $S_{mn}^{(1)}(\eta)$ are orthogonal functions, namely that

$$I_1^{Nn}=\int_{-1}^{+1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta=\begin{cases}0,&n\neq N\\2\sum_{k=0,1}^{\infty}{}'\frac{(d_k^{oN})^2}{2k+1},&n=N.\end{cases}$$

$$(50)$$

The other integrals are much more difficult to handle. The method that is employed to evaluate the remaining integrals is to use (5) to express $S_{mn}^{(1)}(\eta)$ and its derivatives in terms of $P_{m+k}^{m}(\eta)$ and its derivatives. This method involves the manipulation of double series and the evaluation of integrals involving $P_{m+k}^{m}(\eta)$ and its derivatives, but the work is straightforward. Only the results are included, and these are listed below.

$$I_2^{Nn}=\int_{-1}^{+1}\frac{\eta}{\sqrt{1-\eta^2}}S_{1n}^{(1)}(\eta)S_{oN}^{(1)}(\eta)\,d\eta.\quad(51)$$

For n even/odd and N even/odd this integral is zero. For n odd and N even one obtains, since $d_k^{mn}=0$ for $k+n$ odd,

$$I_2^{Nn}=2d_o^{oN}(d_1^{1n}+d_3^{1n}+d_5^{1n}+\cdots)$$

$$+\tfrac{4}{5}d_2^{oN}(d_1^{1n}+d_3^{1n}+d_5^{1n}+\cdots)$$

$$+\tfrac{6}{5}d_2^{oN}(d_3^{1n}+d_5^{1n}+d_7^{1n}+\cdots)$$

$$+\tfrac{8}{9}d_4^{oN}(d_3^{1n}+d_5^{1n}+d_7^{1n}+\cdots)$$

$$+\tfrac{10}{9}d_4^{oN}(d_5^{1n}+d_7^{1n}+d_9^{1n}+\cdots)\quad(52)$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

For n even and N odd the result is

$$-I_2^{Nn}=\tfrac{2}{3}d_1^{oN}(d_0^{1n}+d_2^{1n}+d_4^{1n}+\cdots)$$

$$+\tfrac{4}{3}d_1^{oN}(d_2^{1n}+d_4^{1n}+d_6^{1n}+\cdots)$$

$$+\tfrac{6}{7}d_3^{oN}(d_2^{1n}+d_4^{1n}+d_6^{1n}+\cdots)$$

$$+\tfrac{8}{7}d_3^{oN}(d_4^{1n}+d_6^{1n}+d_8^{1n}+\cdots)$$

$$+\tfrac{10}{11}d_5^{oN}(d_4^{1n}+d_6^{1n}+d_8^{1n}+\cdots)$$

$$+\tfrac{12}{11}d_5^{oN}(d_6^{1n}+d_8^{1n}+d_{10}^{1n}+\cdots)$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$(53)$$

$$I_3^{Nn}=\int_{-1}^{+1}\eta S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)\,d\eta.\quad(54)$$

The integral I_3^{Nn} is zero for n even/odd and N even./odd For n even/odd and N odd/even we find

$$I_3^{Nn}=\sum_{k=1}^{\infty}{}'\Big[\frac{k+1}{k(k+2)}\Big]\Big[d_{\frac{k-1}{2}}^{on}\cdot d_{\frac{k+1}{2}}^{oN}+d_{\frac{k+1}{2}}^{on}\cdot d_{\frac{k-1}{2}}^{oN}\Big]$$

$$(55)$$

$$I_4^{Nn}=\int_{-1}^{+1}(1-\eta^2)\Big[\frac{d}{d\eta}S_{on}^{(1)}(\eta)\Big]S_{oN}^{(1)}(\eta)\,d\eta.\quad(56)$$

The integral I_4^{Nn} also is zero for n even/odd and N even/odd. For n even/odd and N odd/even one obtains

$$I_4^{Nn}=\sum_{k=1}^{\infty}{}'\frac{k+1}{2k(k+2)}\Big[(k+3)\,d_{\frac{k+1}{2}}^{on}\,d_{\frac{k-1}{2}}^{oN}$$

$$-(k-1)\,d_{\frac{k-1}{2}}^{on}\,d_{\frac{k+1}{2}}^{oN}\Big]\quad(57)$$

$$I_5^{Nn}=\int_{-1}^{+1}\sqrt{1-\eta^2}S_{1n}^{(1)}(\eta)S_{oN}^{(1)}(\eta)\,d\eta.\quad(58)$$

This integral is zero if n is even/odd and N is odd/even. For n even/odd and N even/odd one finds

$$I_5^{Nn} = \sum_{k=0}^{\infty'} \left[\frac{k+2}{2(k+1)(k+3)} \right]$$

$$\left[kd^{1n}_{\frac{k}{2}-1} d^{oN}_{\frac{k}{2}+1} - (k+4) \overset{1n}{d}_{\frac{k}{2}} \overset{oN}{d}_{\frac{k}{2}} \right]. \quad (59)$$

$$I_6^{Nn} = \int_{-1}^{+1} \eta \sqrt{1-\eta^2} \left[\frac{d}{d\eta} S_{1n}^{(1)}(\eta) \right] S_{oN}^{(1)}(\eta) \, d\eta. \quad (60)$$

If n is even/odd and N is odd/even, the integral I_6^{Nn} is zero. For n even/odd and N even/odd the integral has the value

$$I_6^{Nn} = \sum_{k=o}^{\infty'} \left[\frac{(k+2)^2}{4(k+1)(k+3)} \right]$$

$$\left[(k+4) d^{1n}_{\frac{k}{2}} d^{oN}_{\frac{k}{2}} + k \, d^{1n}_{\frac{k}{2}-1} d^{oN}_{\frac{k}{2}+1} \right] - \sum_2. \quad (61)$$

For both n and N even, $\sum_2$ is given by

$$\begin{aligned}
\sum_2 = \; & 2d_o^{oN}(\tfrac{1}{3}d_0^{1n} + d_2^{1n} + d_4^{1n} + \cdots) \\
& + \tfrac{4}{5}d_2^{oN}(\tfrac{1}{3}d_0^{1n} + d_2^{1n} + d_4^{1n} + \cdots) \\
& + \tfrac{6}{5}d_2^{oN}(\tfrac{3}{7}d_2^{1n} + d_4^{1n} + d_6^{1n} + \cdots) \\
& + \tfrac{8}{9}d_4^{oN}(\tfrac{3}{7}d_2^{1n} + d_4^{1n} + d_6^{1n} + \cdots) \\
& + \tfrac{10}{9}d_4^{oN}(\tfrac{5}{11}d_4^{1n} + d_6^{1n} + d_8^{1n} + \cdots) \\
& + \tfrac{12}{13}d_6^{oN}(\tfrac{5}{11}d_4^{1n} + d_6^{1n} + d_8^{1n} + \cdots) \\
& \qquad\qquad . \\
& \qquad\qquad . \\
& \qquad\qquad .
\end{aligned}$$

$$(62)$$

and for both n and N odd, $\sum_2$ is given by

$$\begin{aligned}
\sum_2 = \; & \tfrac{2}{3}d_1^{oN}(d_1^{1n} + d_3^{1n} + d_5^{1n} + \cdots) \\
& + \tfrac{4}{3}d_1^{oN}(\tfrac{2}{5}d_1^{1n} + d_3^{1n} + d_5^{1n} + \cdots) \\
& + \tfrac{6}{7}d_3^{oN}(\tfrac{2}{5}d_1^{1n} + d_3^{1n} + d_5^{1n} + \cdots) \\
& + \tfrac{8}{7}d_3^{oN}(\tfrac{4}{9}d_3^{1n} + d_5^{1n} + d_7^{1n} + \cdots) \\
& + \tfrac{10}{11}d_5^{oN}(\tfrac{4}{9}d_3^{1n} + d_5^{1n} + d_7^{1n} + \cdots) \\
& + \tfrac{12}{11}d_5^{oN}(\tfrac{6}{13}d_5^{1n} + d_7^{1n} + d_9^{1n} + \cdots) \\
& + \tfrac{14}{15}d_7^{oN}(\tfrac{6}{13}d_5^{1n} + d_7^{1n} + d_9^{1n} + \cdots) \\
& \qquad\qquad . \\
& \qquad\qquad . \\
& \qquad\qquad .
\end{aligned}$$

$$(63)$$

It is now possible to use (48) and (49) to evaluate the coefficients α_n and β_n. For easier manipulation of the equations, the following substitutions are employed:

$$B_{Nn} = \left(\frac{j}{c\xi_o} \right) A_n (\xi_o^2 - 1)^{1/2} \left[\frac{d}{d\xi} R_{on}^{(1)} \xi \right]_{\xi=\xi_o}$$

$$\int_{-1}^{+1} S_{on}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta, \quad (64)$$

$$C_{Nn} = (\xi_o^2 - 1)^{1/2} \left[\frac{d}{d\xi} R_{on}^{(4)}(\xi) \right]_{\xi=\xi_o}$$

$$\int_{-1}^{+1} S_{on}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta, \quad (65)$$

$$D_{Nn} = -R_{1n}^{(4)}(\xi_o) \int_{-1}^{+1} \frac{\eta}{\sqrt{1-\eta^2}} S_{1n}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta, \quad (66)$$

$$U_{Nn} = \left(\frac{j}{c\xi_o} \right) A_n \Big\{ (\xi_o^2 - 1) \left[\frac{d}{d\xi} R_{on}^{(1)}(\xi) \right]_{\xi=\xi_o}$$

$$\int_{-1}^{+1} \eta S_{on}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta \quad (67)$$

$$+ \xi_o R_{on}^{(1)}(\xi_o) \int_{-1}^{+1} (1 - \eta^2) \left[\frac{d}{d\eta} S_{on}^{(1)}(\eta) \right] S_{oN}^{(1)}(\eta) \, d\eta \Big\},$$

$$V_{Nn} = (\xi_o^2 - 1) \left[\frac{d}{d\xi} R_{on}^{(4)}(\xi) \right]_{\xi=\xi_o}$$

$$\int_{-1}^{+1} \eta S_{on}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta \quad (68)$$

$$+ \xi_o R_{on}^{(4)}(\xi_o) \int_{-1}^{+1} (1 - \eta^2) \left[\frac{d}{d\eta} S_{on}^{(1)}(\eta) \right] S_{oN}^{(1)}(\eta) \, d\eta,$$

$$W_{Nn} = -\xi_o(\xi_o^2 - 1)^{1/2} \left[\frac{d}{d\xi} R_{1n}^{(4)}(\xi) \right]_{\xi=\xi_o}$$

$$\int_{-1}^{+1} \sqrt{1-\eta^2} \, S_{1n}^{(1)}(\eta) S_{oN}^{(1)}(\eta) \, d\eta, \quad (69)$$

$$+ (\xi_o^2 - 1)^{1/2} R_{1n}^{(4)}(\xi_o) \int_{-1}^{+1} \eta \sqrt{1-\eta^2}$$

$$\left[\frac{d}{d\eta} S_{1n}^{(1)}(\eta) \right] S_{oN}^{(1)}(\eta) \, d\eta.$$

Using these symbols in (48) and (49) one obtains

$$\sum_{n=o}^{\infty} (\alpha_n C_{Nn} + \beta_n D_{Nn}) = {}^I E_o a \sum_{n=o}^{\infty} B_{Nn}, \quad (70)$$

$$\sum_{n=o}^{\infty} (\alpha_n V_{Nn} + \beta_n W_{Nn}) = {}^I E_o a \sum_{n=o}^{\infty} U_{Nn}. \quad (71)$$

These are $2N$ equations in $2n$ unknowns, and both N and n go from zero to infinity.

Each term $(\alpha_n{}^x \underline{M}_{on}^{(4)} + \beta_n{}^z \underline{M}_{1n}^{(4)})$ in the expression for ${}^S \underline{E}$, given by (39), represents the reradiation from the spheroid for the forced oscillation of the nth mode. From physical reasoning it is known that the amplitudes of the modes, above a certain order depending upon the ratio F/λ, become less the higher the order of the mode. Hence the coefficients α_n and β_n must approach zero as n increases, and the same fact can be deduced mathematically from the convergence of the series. In obtaining α_n and β_n from (70) and (71), one may expect reasonably accurate values if it is assumed that α_n and β_n are zero for values of n above a given number n'. Then $2n'$ of these equations are employed to find α_n and β_n for values of n up to n'. When these values of α_n and β_n are used in (39), a solution for ${}^S \underline{E}$ is obtained which satisfies the four conditions listed at the beginning of Section I.

It is evident that a large amount of calculating is required in order to obtain numerical results from (39). Some consideration has been given to the possibility of obtaining a simpler solution which would be more

amenable to numerical calculation. To represent $^S\underline{E}$ it would be desirable to have available two sets of vector functions, one with no η-component and the other with no ϕ-component. Then only one set of undetermined coefficients would appear in (33) and only the other set would be present in (34). The process of calculating the coefficients would then be considerably simplified. Simplification would also result if $^S\underline{E}$ were expressed by an equation which contained only one term for each value of n and if this term involved, aside from algebraic functions of η, only $S_{on}^{(1)}(\eta)$ and not its derivatives. In addition it would be helpful if the vector functions used for expressing $^S\underline{E}$ would involve only $S_{on}^{(1)}(\eta)$ with algebraic functions of η identical to those occurring in the expression for $^S\underline{E}$. Then, because of the orthogonality property (50), each coefficient α_n and β_n would be given by a single equation involving no other coefficient. It would be very desirable that the divergence of each of these functions be zero for all values of the eigenvalues n and m; then each term of the series for the scattered wave would be divergenceless with the result that the divergence of the complete expression would be zero, as is necessary. As yet, no solutions of the vector Helmholtz equation have been found which have these desired properties.

VI. Physical Properties of Scattered Wave

Eq. (39), with the appropriate values α_n and β_n as calculated from (70) and (71), is an expression for the electric field vector $^S\underline{E}$ of the scattered wave which is valid for all values of $\eta(-1 \leq \eta \leq +1)$, of $\xi(\xi \geq \xi_o)$, and of ϕ. This means that (39) can be used for calculating $^S\underline{E}$ in the immediate vicinity of the scattering spheroid as well as at great distances from the spheroid.

Usually one is more interested in the behavior of the scattered field at relatively great distances from the spheroid. To deduce this behavior most easily it is well to take the asymptotic form of $^S E$ as $\xi \to \infty$, under the assumption that $c = 2\pi F/\lambda$ does not equal zero. To obtain the asymptotic form of (39) one uses (44) and (45):

$$
^S\underline{E} = \left\{ i_\eta \left[\sum_{n=o}^{\infty} j^{n+2}\alpha_n' S_{on}^{(1)}(\eta) \sin\phi \right] \right.
$$
$$
+ \underline{i}_\phi \left[\sum_{n=o}^{\infty} j\alpha_n'\eta S_{on}^{(1)}(\eta) \cos\phi \right.
$$
$$
\left. \left. + j^{n+3}\beta_n'(1 - \eta^2)^{1/2}S_1\eta^{(1)}(\eta) \cos\phi \right] \right\} \quad (72)
$$
$$
(^I E_o a) \frac{1}{r} e^{-j2\pi r/\lambda}.
$$

Here $\alpha_n' = \alpha_n/^I E_o a$ and $\beta_n' = \beta_n/^I E_o a$. Observe the absence of the $\underline{i}_\xi$ component, which is the longitudinal component for large values of ξ. This, of course, is a necessary and sufficient condition that the wave be a purely transverse wave at a large distance from the spheroid, as it must be.

In order to obtain more information concerning $^S\underline{E}$ it is best to express the electric field vector in terms of components along the rectangular coordinate axes. The resulting expression for $^S\underline{E}$ is

$$
^S\underline{E} = \left\{ \underline{i}_x \left[\sum_{n=o}^{\infty} j^{n+1}\beta_n' \sqrt{1 - \eta^2} S_{1n}^{(1)}(\eta) \sin\phi \cos\phi \right] \right.
$$
$$
+ \underline{i}_y \left[\sum_{n=o}^{\infty} j^n \alpha_n'\eta S_{on}^{(1)}(\eta) + j^{n+3}\beta_n'\sqrt{1 - \eta^2} S_{1n}^{(1)}(\eta) \cos^2\phi \right]
$$
$$
\left. + \underline{i}_z \left[\sum_{n=o}^{\infty} j^{n+2}\alpha_n'\sqrt{1 - \eta^2} S_{on}^{(1)}(\eta) \sin\phi \right] \right\}
$$
$$
\frac{^I E_o a}{r} e^{-j2\pi r/\lambda}. \quad (73)
$$

It will be recalled that the incident wave was taken as moving in the negative z-direction, with the E-vector pointing along the positive y-axis and the H-vector pointing along the positive x-axis. From (73) it can be seen that the back-scattered wave ($\eta = 1$) which is the wave moving in the positive z-direction, has a component of the E-vector along the y-direction only. Thus the back-scattered wave has the same polarization as the incident wave, a well-known property of back-scattered waves from smooth surfaces.

From (73) it may be seen also that the E-vector of scattered radiation being propagated in the xz-plane ($\phi = 0$ or π), has only a y-component, while the E-vector of scattered radiation moving in the yz-plane ($\phi = \pm \pi/2$). has both a y-component and a z-component, except along the y-axis ($\eta = 0$) where the y-component disappears, as it must for a transverse wave.

In order to compute the magnitude of the Poynting vector of the wave scattered by the prolate spheroid, one may use (73) in the following form:
$$
^S\underline{E} = \left\{ \underline{i}_x {}^S T_x(\eta, \phi) + \underline{i}_y {}^S T_y(\eta, \phi) \right.
$$
$$
\left. + \underline{i}_z {}^S T_z(\eta, \phi) \right\} \frac{^I E_o a}{r} e^{-j2\pi r/\lambda} \quad (74)
$$

where $^S T_x$, $^S T_y$, and $^S T_z$ are complex. Now let
$$
^S\underline{T} e^{jt} = \underline{i}_x {}^S T_x + \underline{i}_y {}^S T_y + \underline{i}_z {}^S T_z, \quad (75)
$$
and
$$
^S\underline{T} = \underline{i}_T {}^S T, \quad (76)
$$

where $^S\underline{T}$ and $^S T$ are real.

Then
$$
^S\underline{E} = \underline{i}_T {}^S T \frac{^I E_o a}{r} e^{-j2\pi r/\lambda - t} \quad (77)
$$

From the scattered field determined above, the scattering cross section of the prolate spheroid is readily found. This scattering cross section is defined as the interception cross section σ of an isotropic scatterer which scatters in the direction under consideration the same power density as the prolate spheroid scatters in this direction. The value of σ is given by

$$
\sigma = 4\pi(^S T)^2 a^2. \quad (78)
$$

Solution of Problems in Electromagnetic Wave Theory on a High Speed Digital Calculating Machine

E. K. RITTER†

Summary—This paper contains references to several problems in electromagnetic wave theory which have been solved by numerical methods. In particular, it treats the methods and machines employed by groups at the University of Michigan, Willow Run Research Center, and at the U. S. Naval Proving Ground, Dahlgren, Virginia, in obtaining on a high-speed computing machine a numerical solution for the radar cross section of a prolate spheroid. At the University of Michigan the work was under the direction of K. M. Siegel, while R. A. Niemann was responsible for that done at the Naval Proving Ground.

INTRODUCTION

This paper is concerned primarily with work performed during 1951 and 1952 and reported by Siegel, Gere, Marx, and Sleator,[1] in one of a series of studies in radar cross sections undertaken at the Willow Run Research Center of the University of Michigan's Engineering Research Institute. The first in this series was "Scattering by a Prolate Spheroid," by F. V. Schultz.[2] Most of the others have been directed by Siegel. As will be indicated later, the computation was performed by a group under the direction of R. A. Niemann at the Naval Proving Ground using the large-scale electronic digital computer ADEC (Aiken Dahlgren Electronic Calculator, formerly known as "Mark III").[3]

The author's sole contribution consisted of supplying a large amount of interest and impetus, plus an occasional small "pat on the back" to those who did the work reported herein. He had the pleasure, by the fortuitous circumstance of having changed jobs in the summer of 1952, of being administratively in charge of both groups: those who did the analysis at the University of Michigan and those who performed the machine computation at Dahlgren; thus he was cast in the role "Bugs Bunny" playing tennis with himself!

Although the work of the research teams headed by Siegel and Niemann is probably to date the most significant of any of the problems in electromagnetic wave theory solved by the use of large-scale digital computers, reference must certainly be made, in any discussion of such problems, to the work of other groups; although, most regrettably, it is impossible to present here (because of space, time, and security limitations) anything resembling a complete list.[4]

THE PHYSICAL PROBLEM

The physical problem[5] considered by Schultz, and by Siegel and his colleagues, was that of determining theoretically the behavior of the nose-on radar cross section of a prolate spheroid of fixed dimensions, as the wavelength of the radiation decreased from the (Rayleigh) region, in which the wavelength (λ) is greater than the characteristic dimension (in this case, $2a$, where a is the semimajor axis of the spheroid), to the first maximum in the "resonance" region (in which the wavelength is nearly equal to the characteristic dimension of the reflecting body).

The prolate spheroid was assumed to be perfectly conducting; the incident field was considered to be that due to a plane electromagnetic wave whose direction of propagation is parallel to the major axis of the spheroid; both the incident and the scattered wave were assumed monochromatic with fixed circular frequency ω; and the medium exterior to the scattering body was assumed to be free space, with fixed dielectric constant ϵ and permeability μ, and with zero conductivity.

THE MATHEMATICAL PROBLEM

In general, the solution of the mathematical problem of obtaining an expression for the scattered field from knowledge of the incident field turns upon a consideration of the following boundary value problem:[6]

---

† Georgia Institute of Technology.

[1] K. M. Siegel, B. H. Gere, I. Marx, and F. B. Sleator, "Studies in Radar Cross-Sections XI—The Numerical Determination of the Radar Cross-Section of a Prolate Spheroid," University of Michigan, UMM–126; December, 1953.

[2] University of Michigan, UMM–42; March, 1950.

[3] The results of this computation are shown graphically in Fig. 1 of K. M. Siegel, F. V. Schultz, B. H. Gere, F. B. Sleator, "The Theoretical and Numerical Determination of the Radar Cross-Section of a Prolate Spheroid," pp. 82-91, this issue.

[4] A Table of Values of $\sqrt{\pi\sigma}/\lambda$ prepared by the Scientific Computing Service, Ltd., for P. H. Blundell of the British Ministry of Supply, and published in the classified volume entitled "Reflection Properties of Radar Targets" (Bureau of Ordnance), by M. L. Meeks, N. A. Logan, and C. H. Wilcox; 1954. (Title is unclassified.)

A. N. Lowan, "Tables of Scattering Functions" N. B. S. Appl. Math., Ser. No. 4.

R. O. Gumprecht and C. M. Sliepcevich, "Tables of Light Scattering Functions for Spherical Particles" University of Michigan, 1951.

J. M. Kelso, "Magneto—Ionic Calculations for 150 Kilocycles per Second," Pennsylvania State University, Report No. 24, Ionosphere Research Laboratory; July 20, 1951.

F. Corbato and J. D. Little, "Spheroidal Wave Functions," "Machine Methods of Computation and Numerical Analysis," Mass. Inst. Tech., Quart. Prog. Rep. No. 11, p. 37; December, 1954.

"Eigenvalue Problem for Propagation of Electromagnetic Waves," Project Whirlwind (Mass. Inst. Tech.) Summary Rep. No. 39, September, 1954, and No. 41, March, 1955.

J. L. Uretsky, "Eigenvalues for a Spheroidal Square Well," "Machine Methods of Computation and Numerical Analysis," Mass. Inst. Tech., Quart. Prog. Rep. No. 15, p. 19; March 15, 1955.

[5] *Ibid.*

[6] Siegel, Gere, Marx, and Sleator, op. cit., pp. 4-5.

$\nabla \times \nabla \times \underline{E} - k^2 \underline{E} = 0$, $\nabla \cdot \underline{E} = 0$,

$E_v(u_o, v, w) = -E_v{}'(u_o, v, w)$, $E_w(u_o, v, w) = -E_w{}'(u_o, v, w)$,

$$\lim_{R \to \infty} R[(\underline{N} \cdot \nabla)\underline{E} + ik\underline{E}] = 0.$$

Here $\underline{E}$ represents the scattered electric vector, E' the incident electric vector; the scattering body is represented by the equation $u = u_o$ (a constant) referred to a curvilinear coordinate system $O - uvw$; R is the radius of a sphere with its center in the scattering body and with exterior unit normal $\underline{N}$; $i = \sqrt{-1}$; and $k^2 = \epsilon\mu\omega^2 = 4\pi/\lambda^2$.

In UMM-126,[1] the method of W. W. Hansen[7] has been extended to deal with the boundary value problem just described, yielding a solution in terms of vector wave functions. The radar cross section σ is shown to be

$$\sigma = \lim_{r \to \infty} 4\pi r^2 (E_o)^2/(E_o{}')^2,$$

in which

$$E_o = (E \cdot E^*)^{1/2},$$

E^* is the complex conjugate of E,

$$E_o{}' = (E' \cdot E'^*)^{1/2},$$

and E'^* is the complex conjugate of E'.

By an application of this extension of Hansen's method, Schultz[2] obtained an exact solution to the problem of scattering from a prolate spheroid, $\xi = \xi_o$. Schultz's method employs a prolate spheroidal coordinate system[3] and his solution is expressed by the equations:[1]

$$E = -\frac{1}{F} \sum_{n=0}^{\infty} \left\{ \alpha_n \left[\left(e_\eta \left(\frac{T}{\Delta} \right)^{1/2} S_{on}{}^{(1)} dR_{on}{}^{(4)}/d\xi - e_\xi \left(\frac{\Sigma}{\Delta} \right)^{1/2} dS_{on}{}^{(1)}/d\eta R_{on}{}^{(4)} \right) \sin\phi \right. \right.$$

$$- e_\phi \left(\frac{\eta T}{\Delta} S_{on}{}^{(1)} dR_{on}{}^{(4)}/d\xi + \frac{\xi\Sigma}{\Delta} dS_{on}{}^{(1)}/d\eta R_{on}{}^{(4)} \cos\phi \right]$$

$$- \beta_n \left[\left(e_\eta \frac{\eta}{(\Sigma\Delta)^{1/2}} - e_\xi \frac{\xi}{(T\Delta)^{1/2}} \right) S_{1n}{}^{(1)} R_{1n}{}^{(4)} \sin\phi \right.$$

$$+ e_\phi \frac{(\Sigma T)^{1/2}}{\Delta} \left(\eta dS_{1n}{}^{(1)}/d\eta R_{1n}{}^{(4)} \right.$$

$$\left. \left. \left. - \xi S_{1n}{}^{(1)} dR_{1n}{}^{(4)}/d\xi \right) \cos\phi \right] \right\} \cdot$$

$$\sum_{n=0}^{\infty} [\alpha_n T_o{}^{1/2} S_{on}{}^{(1)}(\eta) dR_{on}{}^{(4)}(\xi_o)/d\xi$$

$$- \beta_n \eta \Sigma^{-1/2} S_{1n}{}^{(1)}(\eta) R_{1n}{}^{(4)}(\xi_o)]$$

$$= \frac{i}{k} E' \sum_{n=0}^{\infty} A_n T_o{}^{1/2} S_{on}{}^{(1)}(\eta) dR_{on}{}^{(1)}(\xi_o)/d\xi$$

$$\sum_{n=0}^{\infty} \left\{ \alpha_n[\eta T_o S_{on}{}^{(1)}(\eta) dR_{on}{}^{(4)}(\xi_o)/d\xi \right.$$

$$+ \xi_o \Sigma dS_{on}{}^{(1)}(\eta)/d\eta R_{on}{}^{(4)}(\xi_o)]$$

$$+ \beta_n (\Sigma T_o)^{1/2} [\eta dS_{1n}{}^{(1)}(\eta)/d\eta R_{1n}{}^{(4)}(\xi_o)$$

$$\left. - \xi_o S_{1n}{}^{(1)}(\eta) dR_{1n}{}^{(4)}(\xi_o)/d\xi] \right\}$$

$$= \frac{i}{k} E' \sum_{n=0}^{\infty} A_n[\eta T_o S_{on}{}^{(1)}(\eta) dR_{on}{}^{(1)}(\xi_o)/d\xi$$

$$+ \xi_o \Sigma dS_{on}{}^{(1)}(\eta)/d\eta R_{on}{}^{(1)}(\xi_o)].$$

$$\sigma = \sigma(1, \phi) = 4\pi a^2 \left| \sum_{n=0}^{\infty} i^n \alpha_n{}' S_{on}{}^{(1)}(1) \right|^2.$$

In these equations the meanings of the symbols are:

F: Semifocal length of spheroid.

ξ: Radial spheroidal variable.

η: Angular spheroidal variable.

ϕ: Azimuthal variable.

T: $T \equiv T(\xi) = \xi^2 - 1$.

∇: $\nabla \equiv \nabla(\xi, \eta) = \xi^2 - \eta^2$.

Σ: $\Sigma \equiv \Sigma(\eta) = 1 - \eta^2$.

$S_{on}{}^{(1)}$: Angular spheroidal function of the first kind.

$R_{on}{}^{(4)}$: Radial spheroidal function of the fourth kind.

$S_{1n}{}^{(1)}$: Angular spheroidal function of the first kind.

$R_{1n}{}^{(4)}$: Radial spheroidal function of the fourth kind.

k: Wave number, $k = \dfrac{2\pi}{\lambda} = \omega\sqrt{\epsilon\mu}$

A_n: Expansion coefficient for incident wave.

E': Amplitude factor of the electric vector E'.

a: Semimajor axis of spheroid, $a = F\xi_o$.

α_n: Expansion coefficient in series for scattered wave.

$\alpha_n{}'$: $\alpha_n{}' = \alpha_n/E'a$.

The first two of these equations represent an exact solution to the vector scattering problem, in the sense that they define, at least in theory, the exact values of all the expansion coefficients in the expression for the scattered field. However, as they stand they are obviously of no value in computing a finite number of the coefficients. Schultz (Ref. 2) suggests the possibility of obtaining simpler expressions for the quantities α_n and β_n by expanding the equations in powers of η and equating the coefficients of like powers, but also points out that the amount of algebra involved in this process is prohibitive. The method developed by Schultz, which has been adopted for the computations to be reported here, is to multiply both members of each equation by $S_{oN}{}^{(1)}(\eta)$ and integrate with respect to η from -1 to 1, for each integer N in the range of the summation index n. The result is an infinite set of linear equations in the infinite set of unknowns α_n and β_n, with constant coefficients. In the notation of Schultz, which will be followed with minor modifications in this section, these equations are

$$\sum_{n=0}^{\infty} (\alpha_n C_{Nn} + \beta_n D_{Nn}) = E'a \sum_{n=0}^{\infty} B_{Nn}, \qquad (1)$$

$$\sum_{n=0}^{\infty} (\alpha_n V_{Nn} + \beta_n W_{Nn}) = E'a \sum_{n=0}^{\infty} U_{Nn} \qquad (2)$$

$(N = 0, 1, 2, \cdots)$, where

[7] W. W. Hansen, "A new type of expansion in radiation problems," *Phy. Rev.*, vol. 47, p. 139; 1935.

$$B_{Nn} = \frac{iA_n}{c\xi_o}(\xi_o^2 - 1)^{1/2}\frac{dR_{on}^{(1)}(\xi_o)}{d\xi}\int_{-1}^{1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta,$$

$$C_{Nn} = (\xi_o^2 - 1)^{1/2}\frac{dR_{on}^{(4)}(\xi_o)}{d\xi}\int_{-1}^{1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta,$$

$$D_{Nn} = -R_{nN}^{(4)}(\xi_o)\int_{-1}^{1}(1-\eta^2)^{1/2}S_{1n}^{(1)}(\eta)S_{oN}^{(1)}(\eta)\,d\eta,$$

$$U_{Nn} = \frac{iA_n}{c\xi_o}\left[(\xi_o^2 - 1)\frac{dR_{on}^{(1)}(\xi_o)}{d\xi}\int_{-1}^{1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta\right.$$
$$\left. + \xi_o R_{on}^{(1)}(\xi_o)\int_{-1}^{1}(1-\eta^2)\frac{dS_{on}^{(1)}(\eta)}{d\eta}S_{oN}^{(1)}(\eta)\,d\eta\right],$$

$$V_{Nn} = (\xi_o^2 - 1)\frac{dR_{on}^{(4)}(\xi_o)}{d\xi}\int_{-1}^{1}S_{on}^{(1)}(\eta)S_{oN}^{(1)}(\eta)d\eta$$
$$ + \xi_o R_{on}^{(4)}(\xi_o)\int_{-1}^{1}(1-\eta^2)\frac{dS_{on}^{(1)}(\eta)}{d\eta}S_{oN}^{(1)}(\eta)d\eta,$$

$$W_{Nn} = -\xi_o(\xi_o^2 - 1)^{1/2}\frac{dR_{1n}^{(4)}(\xi_o)}{d\xi}\int_{-1}^{1}(1-\eta^2)^{1/2}S_{1n}^{(1)}(\eta)$$
$$S_{oN}^{(1)}(\eta)\,d\eta$$
$$ + (\xi_o^2 - 1)^{1/2}R_{1n}^{(4)}(\xi_o)\int_{-1}^{1}(1-\eta^2)^{1/2}$$
$$\frac{dS_{1n}^{(1)}(\eta)}{d\eta}S_{oN}^{(1)}(\eta)\,d\eta.$$

By taking only the first few terms of the series, and the same range of values of N, one may derive a finite system of linear equations whose solution should approximate the exact values. The degree of approximation will of course depend on the order of the system. A fourth-order system was actually used for the computations, the choice being dictated primarily by limitations on machine capacity, time, and money. It was hoped that the degree of approximation obtained in this way would be at least comparable to that afforded by the other approximations inherent in the general method. Also, of course, the problem was simplified in that computations were made only for the case of back-scattering, so that the coefficients β_n drop out of the picture.

Among the approximations made in the computations, the principal one concerns the representations of the spheroidal functions $S_{mn}(\eta)$ and $R_{mn}(\xi)$. Because of the lack of knowledge of the theory of these functions, and because of the limited range of existing tables, it was necessary to represent the functions S_{mn} and R_{mn} by series of associated Legendre functions and of spherical Bessel functions, respectively.

The ADEC

The machine employed for the present computations was the ADEC (Aiken Dahlgren Electronic Calculator, formerly known as "Mark III"). This computing machine was constructed for the Bureau of Ordnance, U.S. Navy, by the Harvard Computation Laboratory, and is now located at the Naval Proving Ground at Dahlgren, Virginia. The ADEC utilizes magnetic drum storage units, and at the time of its selection for this

work had the largest capacity of any digital machine available. Previously, one value of a radar cross section had been obtained at the cost of several man-months of manual computation. To compute radar cross sections with the necessary accuracy for the desired number of values of the parameters, it was necessary to use a machine with large storage capacity and large word length; this was the principal consideration leading to the choice of the ADEC. The difference in operating costs between this machine and faster machines with lower capacities was outweighed by the reduction in computing time made possible by the storage facilities of the ADEC.

These facilities include channels capable of storing 4,350 sixteen-decimal digit numbers, of which 150 are permanently stored constants. In addition, the machine has a capacity of 4,000 three-address instructions, and carries permanently stored routines for elementary operations and computational procedures. In this problem floating arithmetic subroutines were employed using fifteen-decimal-digit coefficients, four-digit exponents and signs. The machine operates with a coded decimal system, in which each decimal digit is represented by four binary digits. Input is supplied on magnetic tapes, and results are printed in decimal numbers by five electric typewriters.

Even with the storage capacity available, the various components of the problem could not be coded for simultaneous computation. It was found necessary to split the problem into five individually coded parts[8] described in some detail in the following section.

A considerable amount of the numerical analysis, development of special machine-computational methods, and all of the programming and coding of the problem as well as the operation of the machine was done at the Naval Proving Ground, Dahlgren, Virginia, under the direction of R. A. Niemann, then Chief of the Computation Division of the Computation and Ballistics Department. The preparation of the machine program was done by G. H. Gleissner, D. F. Eliezer, Karl Kozarsky, and A. M. Fleishman.

The Machine Computations

A schematic diagram of the subdivision of the problem, indicating the interdependence of the various intermediate steps is given in Fig. 1. The distribution of these steps among the five runs is shown by listing in a separate column the quantities computed in each run. The order of the computations proceeds from left to right. The order in which items appear in their column is not significant.

Some idea of the magnitude of the computational effort required may be gained from a partial listing of the expressions contained in the blocks of Fig. 1. For example:[9]

[8] Referred to as Runs 1, 2, 3, 4, 5 in the diagram which appears in the next section.
[9] For definitions of symbols in these equations not found in this
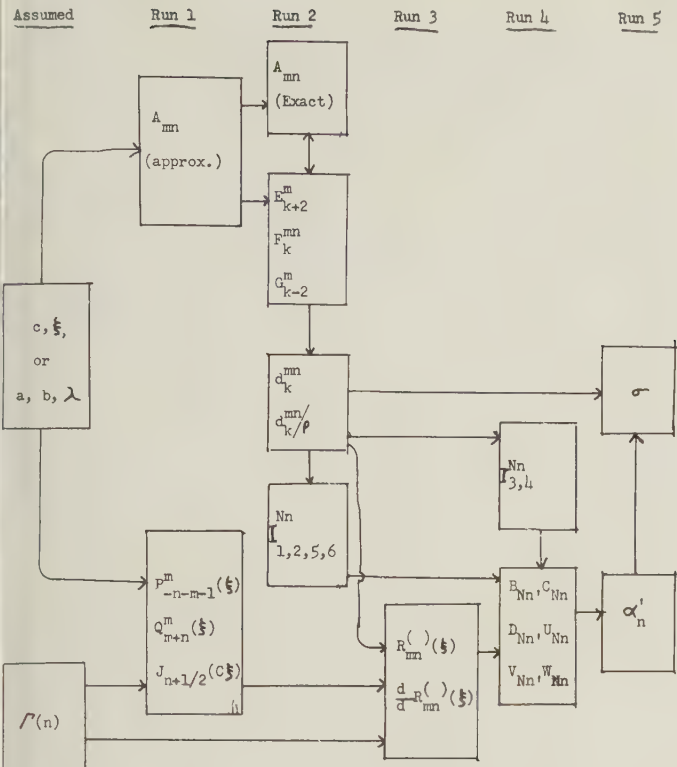
Fig. 1—Logical and sequential structure of computation.

$$E_{k+2}{}^m d_{k+2}{}^{mn} + F_k{}^{mn} d_k{}^{mn} + G_{k-2}{}^m d_{k-2}{}^{mn} = 0 \qquad (4)$$

$$E_{k+2}{}^m = \frac{(2m+k+2)(2m+k+1)}{(2m+2k+3)(2m+2k+5)} \qquad (5)$$

$$F_k{}^{mn} = \frac{(m+k)(m+k+1) + A_{mn}}{c^2}$$
$$+ \frac{2(m+k)(m+k+1) - 2m - 1}{(2m+2k-1)(2m+2k+3)} \qquad (6)$$

$$G_{k-2}{}^m = \frac{k(k-1)}{(2m+2k-3)(2m+2k-1)} \qquad (7)$$

$$R_{mn}{}^{(2)}(\xi) = K(c,m,n)\left[\sum_{k=-\infty}^{-2m-1} \mu_{n,k} d_k{}^{mn} P_{-n-1}(\xi)/\rho \right.$$
$$\left. + \sum_{k=2m}^{\infty} \mu_{n,k} d_k{}^{mn} Q_n(\xi)\right].$$

$$I_6{}^{Nn} = \int_{-1}^{1} \eta(1-\eta^2)^{1/2} S_{oN}{}^{(1)}(\eta) \frac{d}{d\eta} S_{1n}{}^{(1)}(\eta)\, d\eta$$

$$= 2\mu_{N,n}\left\{\sum_{k=0}^{\infty} \frac{(k+1)^2}{(2k+1)(2k+3)}\left[(k+2)d_k{}^{1n}d_k{}^{oN}\right.\right.$$
$$+ kd_{k-1}{}^{1n}d_{k+1}{}^{oN}\Big] - \sum_{k=0}^{\infty} \frac{\mu_{n,k}}{2k+1} d_k{}^{oN}\left[\frac{k(k-1)}{2k-1}d_{k-2}{}^{1n}\right.$$
$$\left. + \frac{(k+1)^2}{2k+3}d_k{}^{1n} + k\, d_k{}^{1n}\right] - \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \mu_{n,k}\mu_{n,j} d_k{}^{oN} d_j{}^{1n}\Bigg\}.$$

$$(\xi^2 - 1)\alpha_{mn}(\xi)dR_{mn}{}^{(j)}/d\xi + \beta_{mn}(\xi)R_{mn}{}^{(j)} = R_{m,n-1}{}^{(j)}$$

$$\alpha_{mn} = \int_{-1}^{1} \frac{\tau}{\xi^2 - \tau^2} S_{mn}{}^{(1)}(\tau)S_{m,n-1}{}^{(1)}(\tau)\, d\tau/$$
$$c\int_{-1}^{1} \tau S_{mn}{}^{(1)}(\tau)S_{m,n-1}{}^{(1)}(\tau)\, d\tau$$
$$= c\left[R_{m,n-1}{}^{(1)}R_{mn}{}^{(2)} - R_{mn}{}^{(1)}R_{m,m-1}{}^{(2)}\right];$$

$$\beta_{mn} = \xi\int_{-1}^{1} \frac{1-\tau^2}{\xi^2 - \tau^2} dS_{mn}{}^{(1)}(\tau)/d\tau S_{m,n-1}{}^{(1)}(\tau)\, d\tau$$
$$\div c\int_{-1}^{1} \tau S_{mn}{}^{(1)}(\tau)S_{m,n-1}{}^{(1)}(\tau)\, d\tau = c(\xi^2 - 1)[R_{m,n-1}{}^{(1)}$$
$$R_{mn}{}^{(2)\prime} - R_{mn}{}^{(1)}{}^1 R_{m}{}^7{}_{n-1}{}^{(2)}] = [R_{m,n-1}{}^{(1)}R_{mn}{}^{(2)\prime} -$$
$$R_{mn}{}^{(1)\prime}R_{m,n-1}{}^{(2)}]/[R_{mn}{}^{(1)}R_{mn}{}^{(2)\prime} - R_{mn}{}^{(1)\prime}R_{mn}{}^{(2)}]$$

$$c < 5: \qquad A_{mn} \approx \sum_{k=0}^{4} g_k{}^{mn}c^{2k}, \qquad (3)$$

where

$$g_o{}^{mn} = -(n+m)(n+m+1),$$

$$g_1{}^{mn} = \frac{1}{2}\left[\frac{(2m-1)(2m+1)}{(2n+2m-1)(2n+2m+3)} - 1\right]$$

$$g_2{}^{mn}$$
$$= \frac{1}{2}\left[\frac{(n+1)(n+2)(n+2m+1)(n+2m+2)}{(2n+2m+1)(2n+2m+3)^3(2n+2m+5)}\right.$$
$$\left. - \frac{(n+2m)(n-1)n(n+2m-1)}{(2n+2m-3)(2n+2m-1)^3(2n+2m+1)}\right],$$

$$g_3{}^{mn}$$
$$= (4m^2-1)\left[\frac{(n-1)n(n+2m-1)(n+2m)}{(2n+2m-5)(2n+2m-3)(2n+2m-1)^5(2n+2m+1)(2n+2m+3)}\right.$$
$$\left. - \frac{(n+1)(n+2)(n+2m+1)(n+2m+2)}{(2n+2m-1)(2n+2m+1)(2n+2m+3)^5(2n+2m+5)(2n+2m+7)}\right]$$

$$g_4{}^{mn} = 2(4m^2-1)^2\left[\frac{(n+1)(n+2)(n+2m+1)(n+2m+2)}{(2n+2m-1)^2(2n+2m+1)(2n+2m+3)^7(2n+2m+5)(2n+2m+7)^2}\right.$$
$$\left. - \frac{(n-1)n(n+2m-1)(n+2m)}{(2n+2m-5)^2(2n+2m-3)(2n+2m-1)^7(2n+2m+1)(2n+2m+3)^2}\right]$$
$$+ 2^{-4}\left[\frac{(n+1)(n+2)(n+3)(n+4)(n+2m+1)(n+2m+2)(n+2m+3)(n+2m+4)}{(2n+2m+1)(2n+2m+3)^4(2n+2m+5)^3(2n+2m+7)^2(2n+2m+9)}\right.$$
$$\left. - \frac{(n-3)(n-2)(n-1)n(n+2m-3)(n+2m-2)(n+2m-1)(n+2m)}{(2n+2m-7)(2n+2m+5)^2(2n+2m-3)^3(2n+2m-1)^4(2n+2m+1)}\right]$$
$$+ 2^{-3}\left[\frac{(n-1)^2n^2(n+2m-1)^2(n+2m)^2}{(2n+2m-3)^2(2n+2m-1)^7(2n+2m+1)^2} - \frac{(n+1)^2(n+2)^2(n+2m+1)^2(n+2m+2)^2}{(2n+2m+1)^2(2n+2m+3)^7(2n+2m+5)^2}\right]$$
$$- \frac{1}{2}\left[\frac{(n-1)n(n+1)(n+2)(n+2m-1)(n+2m)(n+2m+1)(n+2m+2)}{(2n+2m-3)(2n+2m-1)^4(2n+2m+1)^2(2n+2m+3)^4(2n+2m+5)}\right].$$

$$W_{Nn} = -\xi_o(\xi_o^2 - 1)\frac{d}{d\xi}R_{1n}^{(4)}(\xi_o)I_5^{Nn}$$
$$+ (\xi_o^2 - 1)^{1/2}R_{1n}^{(4)}(\xi_o)I_6^{Nn}.$$

$$\alpha_o' = \frac{1}{G}\begin{vmatrix} B_{oo} & O & D_{01} & D_{03} \\ & & & \\ B_{22} & C_{22} & D_{21} & D_{23} \\ U_{10} + U_{12} & V_{12} & W_{11} & W_{13} \\ U_{30} + U_{32} & V_{32} & W_{31} & W_{33} \end{vmatrix}$$

$$G = \begin{vmatrix} C_{oo} & O & D_{01} & D_{03} \\ O & C_{22} & D_{21} & D_{23} \\ V_{10} & V_{12} & W_{11} & W_{13} \\ V_{30} & V_{32} & W_{31} & W_{33} \end{vmatrix}$$

$$\sigma = 4\pi a^2 \left| \sum_{n=o}^{\infty} i^n \alpha_n' \sum_{k=O}^{\infty} d_k^{on} \right|^2.$$

In the course of manual computations it was discovered that for most values of c used, and particularly for those near $c = 5$, the number of coefficients given in (3) and in its counterpart for $c > 5$ was not enough to determine the separation constants to the required degree of approximation, while the formulas for additional coefficients were prohibitively complicated. A small error in the value of A_{mn} led to errors in the recurrence formulas for the spheroidal coefficients d_k^{mn} of the order of magnitude of the quantities computed. For this reason an iterative procedure was devised that served to refine the values of A_{mn} sufficiently in all cases computed. The procedure is outlined as applied for even values of k; for odd k there are minor modifications:

1. The quantities E_{k+2}^m, F_k^{mn}, and G_{k-2}^m are computed as in (5), (6), and (7), with the value $A_{mn} = A_{mn}^{(1)}$ determined by the series (3) or its counterpart for $c > 5$.

2. The value of $K_k^{mn} = d_{k+2}^{mn}/d_k^{mn}$ is found for $k = 14$ from the formula $K_{14}^{mn} = -G_{14}^{mn}/F_{16}^{mn}$, obtained from (4) with $k = 16$ if d_{18}^{mn}/d_{14}^{mn} is replaced by its limit zero.

3. The values of K_k^{mn} for $k = 12, 10, \cdots, 0$ are found successively from the recurrence formula (4) used in the form

$$1/K_{k-2}^{mn} = -(F_k^{mn} + E_{k+2}^m K_k^{mn})/G_{k-2}^m.$$

4. The quantity $F_o^{mn} + E_2^m K_o^{mn} = -G_{-2}^m/K_{-2}^{mn}$ is computed. As follows from (7), $G_{-2}^m = 0$ is the exact value that should be obtained, and so $|G_{-2}^m/K_{-2}^{mn}|$ is a measure of the error in the first approximation $A_{mn}^{(1)}$ to the separation constant. If the error is not considered excessive, the values K_k^{mn} obtained in Step 3 may be used to compute the spheroidal coefficients d_k^{mn}.

5. If the error exceeds the bounds of required accuracy, a second approximation $A_{mn} = A_{mn}^{(2)}$ to the separation constant is derived from the formula (4) for $k = 0$, which may be written

$K_o^{mn} = -[c^{-2}(m^2 + m + A_{mn}) + 1/(2m + 3)]/E_2^m$; the value K_o^{mn} used is the one computed in Step 4. An average of the two approximations $A_{mn}^{(1)}$ and $A_{mn}^{(2)}$ is then used to begin the process again at Step 1.

If the procedure is stable, the error in Step 4 will decrease to within the desired range of accuracy after a sufficient number of iterations. If the error does not decrease, the method is inapplicable; in the present computations, this difficulty did not arise.

The procedure described above is similar to a method developed by C. J. Bouwkamp.[10] The method of Bouwkamp is theoretically superior in improving the approximations to A_{mn}, converging more rapidly and for a wider range of parameter values. However, it is also much more difficult to program for a digital machine like the ADEC, and this fact led to its rejection in favor of the procedure outlined above.

QUANTITIES TABULATED IN ADEC OUTPUT

The values of $2\pi a/\lambda$ for which the radar cross section was computed are listed below:
.094, .105, .157, .236, .262, .314, .377, .524, .754, .785, .942, 1.26, 1.57, 1.89, 2.10, 2.51, 3.14, 3.77, 4.71, 5.89, 6.28, 10.5.

In the computations a system of four equations in four unknowns replaced the infinite system (1), (2). For purposes of comparison, the system of three equations in three unknowns was computed for the value $\frac{2\pi a}{\lambda} = 1.89$.

For each value of $\frac{2\pi a}{\lambda}$ listed, the following quantities are recorded:

1. Associated Legendre functions $P_{-n-m-1}^m(\xi)$ and $Q_{m+n}^m(\xi)$ and their derivatives with respect to ξ,

for $m = 0, 1$,

$$n = -2, \ldots, -16.$$

2. Bessel functions $J_{n+1/2}(c\xi)$, for $n = 1, 2, \cdots, 17$.

3. Separation constants A_{mn}, for $m = 0, 1$; $n = 0, 1, 2, 3$.

4. Spheroidal coefficients d_k^{mn} (or d_k^{mn}/ρ in range of $k < 0$, where $d_k^{mn} = 0$),
for $m = 0, 1$,
  $n = 0, 1, 2, 3$,
  $k =$ all necessary values in range -16 to $+16$.

5. Radial spheroidal functions $R_{mn}^{(1)}(\xi)$ and $R_{mn}^{(2)}(\xi)$ and derivatives of these with respect to ξ,

for $\xi = 1.005$
$$m = 0, 1$$
$$n = 0, 1, 2, 3.$$

---

[10] C. J. Bouwkamp, "On spheroidal wave functions of order zero," *Jour. Math. Phys.*, vol. 26, p. 79; 1947.

6. Boundary integrals I_k^{Nn}
    for $k = 1, 2, 3, 4, 5, 6,$

all combinations of N and n in range

$$N = 0, 1, 2, 3,$$
$$n = 0, 1, 2, 3.$$

7. Radar cross section σ.

With the exception of σ, all quantities are given to 15 significant figures. The values of σ are rounded off to 5 significant figures.

Conclusion

The accuracy required in the computations described in this paper (i.e., the large number of decimal places necessary), as well as the need for a large amount of storage capacity for intermediate computations, dictated the choice of the ADEC for the purpose. In 1952 ADEC was still undergoing its "shake-down cruise." Early in 1953 engineering improvements were installed which would materially reduce the time required to solve this problem, were additional solutions desired. For example, approximately *twenty-four hours* were required to read into the machine and check *each* of the five parts of the program. This can now be done in half an hour. The efficiency of operation of the ADEC (roughly the percentage of "good" machine time available) has increased from about 50 per cent in 1952 to about 85 per cent in 1955. If the problem were to be repeated on ADEC in 1955, it could probably be solved in half the time which was required in 1952.

Moreover, several other machines are now available which exceed the ADEC in speed and memory capacity, although most of the newer machines do not equal the ADEC in number of decimal digits available without resort to "multiple precision" techniques. The NORC (Naval Ordnance Research Calculator), recently installed at the Naval Proving Ground, has a word length of sixteen decimal digits, built-in checking, and floating-decimal operation. A few months ago it solved in forty-two *minutes* a problem which had required three hundred *hours* on ADEC last year.

In 1952 no universities and few government agencies had computing machines capable economically of solving the problem of obtaining the radar cross-section of a prolate spheroid. Today the Georgia Institute of Technology, the University of Michigan, the University of Illinois, and several other institutions of higher learning in this country, as well as a number of industrial laboratories, have large-scale electronic machines which could handle this problem and many others at least as difficult.

Edge Currents in Diffraction Theory

P. C. CLEMMOW†

Summary—A comparatively simple method for obtaining an asymptotic approximation to the electromagnetic field diffracted by a large aperture in a perfectly conducting, infinitely thin, plane screen is suggested. The method is based on two assumptions: first, that in some regions the scattered field is nearly the same as the field that would be generated by certain currents located on the edge of the aperture; secondly, that at any point on the edge of the aperture these currents are nearly the same as the corresponding currents for a half-plane lying in the plane of the diffracting screen, the straight edge of which is locally coincident with the edge of the aperture. In the crudest approximation the calculation is made on the basis that the half-planes are excited by the incident field alone; higher order approximations arise from a consideration of the interaction between the different parts of the edge of the aperture.

Applications of the method to the cases of a plane wave normally incident on (1) a slit of infinite length with parallel straight edges, and (2) a circular aperture are considered. In the former case several terms of the asymptotic development of the transmission cross section in inverse powers of the slit width are given; in the latter case the aperture and axial fields based on the zero-order approximation which neglects interaction are compared with experimental data published by various authors and with some rigorous calculations of Andrejewski.

1. INTRODUCTION

The problem considered is the diffraction of a given electromagnetic field by an aperture in a perfectly conducting, infinitely thin, plane screen. As is well known, the only case with a reasonably simple exact solution is that of the half-plane with a straight edge. In other cases most methods of solution have yielded numerical results only when the relevant dimensions of the aperture are less than or comparable with a wavelength. Relatively little attention has been paid to the interesting problem when the relevant dimensions of the aperture are large compared with the wavelength, as may be judged, for example, from Bouwkamp's[1] exhaustive review of recent work on diffraction theory. Indeed it appears that only in the last few years has it become widely appreciated that in some particulars the predictions of geometrical optics are quite inadequate however large the aperture.

In the present paper consideration is given to an approximate method which would appear to yield fairly readily, for rather general shapes of aperture, reasonably accurate estimations of the field in certain regions, including the aperture itself. The method is akin to that of Braunbek[2] and Frahn,[3] but perhaps somewhat simpler to work out; it probably gives the early terms in an asymptotic development of the exact solution in inverse powers of the relevant dimensions of the aperture.

Very early in the history of the diffraction problem both experimental and theoretical work suggested that in some respects the diffracted field could be regarded as emanating from a source located on the diffracting edge. In terms of an aperture in a perfectly conducting screen this feature is more precisely expressed by the statement that, at some points, the scattered field which is radiated by the *surface* currents induced in all parts of the screen is nearly the same as that which would be radiated by certain *line* currents located on the edge of the aperture. In the particular case of a half-plane with a straight edge the simple exact solution makes manifest the way in which this arises; namely, by virtue of the fact that the terms of the *asymptotic* expansion of the solution in inverse powers of distance from the edge can be related to line currents located on the edge.

Furthermore, at points on the edge of the aperture where the radius of curvature is large compared to the wavelength, it seems reasonable to assume that the edge currents are the same as those pertaining to the half-plane, lying in the plane of the screen, the straight edge of which is locally coincident with the edge of the aperture.

The proposed method, then, may be stated briefly as follows: by using the known half-plane results, edge currents are derived from which the scattered field at certain points may be calculated.

On this basis the crudest approximation is obtained by neglecting the interaction between different parts of the edge; or, in other words, by taking the edge currents at each point as those of the appropriate half-planes excited by the incident field alone. Better approximations would take account, in some way, of the fact that each half-plane is also excited by the field of the edge currents at other points of the aperture.

The points at which the scattered field can validly be calculated in the proposed manner depend on the nature of the incident field. Roughly speaking, on the side of the screen remote from the source, they lie in the illuminated region not too close to the boundary of the "shadow" cast by the screen; this region includes points in the aperture itself, provided they are not too close to the edge.

The method evidently gives the scattered field in terms of line integrals. In this respect it yields formulas similar to those proposed by Andrews,[4] though there are significant differences in the integrands.

These and other details of the method, together with its application to the circular aperture, will be reported elsewhere by Mr. R. F. Millar. In the present paper some results of the application of the method to the

† Cambridge University, Cambridge, England.
[1] C. J. Bouwkamp, Rep. Prog. Phys., vol. 17, p. 35; 1954.
[2] W. Braunbek, *Z. Phys.*, vol. 127, pp. 381, 405; 1950.
[3] W. Frahn, Thesis, Aachen; 1951.

[4] C. L. Andrews, *Jour. Appl. Phys.*, vol. 21, p. 761 ; 1950.

slit are given in Section 2, attention being largely confined to the transmission cross section for a normally incident plane wave; and in Section 3 some of the aperture and axial fields obtained by Millar for the circular aperture, neglecting interaction, are compared with experimental values and with some rigorous calculations of Andrejewski.[5]

It should perhaps be stressed that the aim of the method is to obtain some features of the behavior of the diffracted field with reasonable accuracy and relative ease. In the particularly simple case of the slit it will indeed yield indefinitely, with additional labor, additional terms of the asymptotic expansion of the field in inverse powers of the slit width; but for more elaborate apertures it is unlikely to be feasible to go beyond the first term or so. In the spirit of the method some matters of rigor are left for subsequent examination, if any.

2. THE SLIT

2.1 *Specification of the Problem*

In this section is considered the two-dimensional problem of a plane wave incident normally on a slit of width d, with infinitely long parallel straight edges, cut in a perfectly conducting, infinitely thin, plane sheet. In a system of rectangular cartesian axes the conducting sheet occupies the whole plane $y = 0$ except for the region of the slit which is $-\frac{1}{2}d < x < \frac{1}{2}d$, and everything is independent of z.

For brevity of description in the sequel the half-plane $y = 0$, $x > \frac{1}{2}d$ is called $P+$, and the half-plane $y = 0$, $x < -\frac{1}{2}d$ is called $P-$. It is also convenient to introduce polar co-ordinates r_1, θ_1 and r_2, θ_2 measured from the edges of $P+$ and $P-$ respectively, and r to denote distance from the z-axis (see Fig. 1).
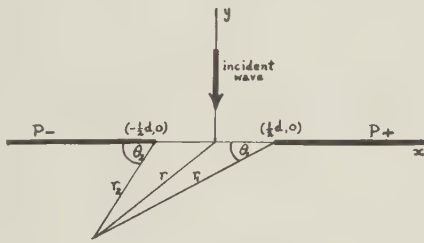
Fig. 1—The notation and configuration for the slit problem.

The two fundamental polarizations are treated separately. The plane wave is taken to be incident from the region $y > 0$, so in the case of *E-polarization* it is given by

$$E_z{}^i = e^{iky}, \tag{1}$$

and in the case of *H-polarization by*

$$H_z{}^i = e^{iky}, \tag{2}$$

the time factor $e^{i\omega t}$ being left understood ($k = \omega/c$).

In what follows the discussion is explicitly confined to the derivation of an approximation, valid when $kd >> 1$,

to the transmission cross section of the slit, or, what is virtually the same thing, the radiation term of the field at $x = 0$ as $y \rightarrow -\infty$. However, it will be apparent that, to the same degree of approximation, the complete field at any point could be obtained if desired.

2.2 *The Appropriate Version of the Cross Section Theorem*

As $r \rightarrow \infty$ for $y < 0$ the leading term in the expression for total H_z in the case of H-polarization and total E_z in the case of E-polarization is some function of angle times $e^{-ikr}/\sqrt{(kr)}$. The version of the cross section theorem (see, for example, Van de Hulst,[6] Levine and Schwinger[7]) appropriate to the present problem states that if on the negative y-axis [the "forward" direction for the incident waves (1) and (2)] these leading terms are written in the form

$$\frac{e^{\frac{1}{4}i\pi}}{\sqrt{(2\pi)}} k \frac{e^{-ikr}}{\sqrt{(kr)}} T, \tag{3}$$

then the *transmission cross section* (per unit length in the z-direction) *is the real part of* T.

2.3 *The Zero-Order Approximation Neglecting Interaction*

In this approximation the scattered field is calculated from the current that would be induced in $P+$ by the incident field alone were $P-$ absent, together with the field that would be induced in $P-$ by the incident field alone were $P+$ absent. That is, the interaction between $P+$ and $P-$ is completely neglected. Since this interaction can roughly be described as the excitation of $P+$ by a cylindrical wave emanating from the edge of $P-$, and vice versa, it is to be expected that the omitted terms are down at least by the order of a factor $(kd)^{-1/2}$. That this is indeed the case is shown explicitly later.

Neglecting interaction, then, the scattered field can be at once written down, from the well-known solution to the half-plane problem, in terms of the complex Fresnel integral

$$F(a) = \int_a^\infty e^{-i\lambda^2} d\lambda. \tag{4}$$

Since our object is to obtain the transmission coefficient, only points on the negative y-axis need be considered. By symmetry the scattered field there receives a contribution arising from the current in $P+$ which is identical with that arising from the current in $P-$, so that, denoting it by a suffix s, for E-polarization

$$E_z{}^s = -\frac{2e^{1/4i\pi}}{\sqrt{\pi}} \{e^{iky}F[\sqrt{(2kr_1)} \sin (\tfrac{1}{4}\pi - \tfrac{1}{2}\theta_1)]$$
$$+ e^{-iky}F[\sqrt{(2kr_1)} \sin (\tfrac{1}{4}\pi + \tfrac{1}{2}\theta_1)]\}, \tag{5}$$

for H-polarization

$$H_z{}^s = -\frac{2e^{1/4i\pi}}{\sqrt{\pi}} \{e^{iky}F[\sqrt{(2kr_1)} \sin (\tfrac{1}{4}\pi - \tfrac{1}{2}\theta_1)]$$
$$- e^{-iky}F[\sqrt{(2kr_1)} \sin (\tfrac{1}{4}\pi + \tfrac{1}{2}\theta_1)]\}. \tag{6}$$

[5] W. Andrejewski, *Z. Angew. Phys.*, vol. 5, p. 178 ; 1953.

[6] H. C. Van de Hulst, *Physica*, vol. 15, p. 740 ; 1949.
[7] H. Levine and J. Schwinger, *Phys. Rev.*, vol. 74, p. 958 ; 1948. *Comm. Pure App. Math.*, vol. 3, p. 355 ; 1950.

The *total* field for the respective polarizations is obtained by adding to (5) the incident field (1), and to (6) the incident field (2).

Now the argument of the first Fresnel integral in (5) and (6) is, in fact, just

$$\sqrt{\{k(r_1 - r)\}} = \frac{kd}{2\sqrt{(2kr)}} + O\{(kr)^{-3/2}\} \text{ as } r \to \infty, \quad (7)$$

and that of the second Fresnel integral in (5) and (6) is

$$\sqrt{\{k(r_1 + r)\}} = \sqrt{(2kr)} + O\{(kr)^{-1/2}\} \text{ as } r \to \infty, \quad (8)$$

where r is conveniently written for $-y$ on the negative y-axis. Furthermore,

$$F(a) = \sqrt{\pi} e^{-1/4 i\pi} - a + O(a^2), \text{ for small } a, \quad (9)$$

$$e^{ia^2} F(a) = \frac{1}{2ia} + O(a^{-3}), \text{ for large } a; \quad (10)$$

(9) coming from the ascending power series and (10) from the asymptotic development.

Hence the radiation part of the total field in the forward direction along the negative y-axis is readily seen to be,

for E-polarization

$$E_z \sim \frac{e^{1/4 i\pi}}{\sqrt{(2\pi)}} kd \left(1 + \frac{i}{kd}\right) \frac{e^{-ikr}}{\sqrt{(kr)}}, \quad (11)$$

for H-polarization

$$H_z \sim \frac{e^{1/4 i\pi}}{\sqrt{(2\pi)}} kd \left(1 - \frac{i}{kd}\right) \frac{e^{-ikr}}{\sqrt{(kr)}}. \quad (12)$$

With the neglect of interaction between P+ and P− as the initial basis of the calculation, (11) and (12) have been derived without further approximation. Appealing to (3) they yield

$$\text{transmission cross section} = d \quad (13)$$

for each polarization. The theory in its crudest form therefore gives only the geometrical optics transmission cross section, though it would evidently lead to correction terms to the field components which would be significant, for example, in the aperture plane.

It should be remarked that, in this first stage of the calculation of the transmission cross section, the edge current concept is not applicable and of necessity use has been made, implicitly at least, of the complete half-plane current distribution. This is because, for the incident fields (1) and (2), the regions in the neighbourhood of the shadow boundaries within which the scattered field cannot be represented by edge waves include all points on the negative y-axis for which kr is larger than some value determined by kd. The corresponding feature of the mathematics is the use of the initial terms of the ascending power series rather than those of the asymptotic development for the leading Fresnel integral in (5) and (6).

The effect of interaction between P+ and P− is considered in the following sections. The method from now on makes use wholly of the edge wave representation, and indeed it will become clear that this representation goes hand in hand with the asymptotic development of the solution in inverse powers of kd.

2.4 A Better Approximation for the Case of E-Polarization

It is not difficult to appreciate that the interaction between P+ and P− is stronger for E-polarization than for H-polarization. The former is considered here and in Section 2.5, the latter in Section 2.6.

In Section 2.3 the analysis was based on the use of what may be called the undisturbed half-plane currents in the plates P+ and P−. As a first step in allowing for interaction the excitation of P− by the field of the undisturbed half-plane current in P+ (and vice versa) must be worked out to the appropriate degree of approximation.

This excitation of the plates results in additional induced currents which in turn play two roles: for the field radiated by that in one plate both contributes to the field in the forward direction, thus giving the first correction to the transmission cross section of geometrical optics, and also induces further currents in the other plate. And so on.

Throughout the analysis the field scattered by P+, say, can be represented on P− as an edge wave. More precisely, it can be represented asymptotically as a series of cylindrical waves the amplitudes of which successively decrease with distance from the edge of P+ as $(kr_1)^{-1/2}$, $(kr_1)^{-3/2}$, $(kr_1)^{-5/2}$, $\cdots$. The respective fields radiated by the currents thus induced in P− therefore contain factors $(kd)^{-1/2}$, $(kd)^{-3/2}$, $(kd)^{-5/2}$, $\cdots$. It follows that, at each stage in the process, account need only be taken of a finite number of the cylindrical waves, the contribution of the predominant one being down by a factor of the order $(kd)^{-1/2}$ on that of the previous stage. Both this number, which thus decreases at each stage, and the number of stages are, of course, determined by the particular inverse power of kd to which the calculation is being carried out.

By this method, then, any one term in the asymptotic development of the field in inverse powers of kd is made up of all contributions having the appropriate combination of order of edge wave and number of times the excitation has passed back and forth across the slit. The derivation of the first two correction terms to the geometrical optics transmission cross section is very easy and is given in this section. The derivation of subsequent terms becomes progressively more complicated, and only two further ones have so far been obtained; these are given in Section 2.5.

The field in $y = 0$, $x < \frac{1}{2} d$ scattered by P+ when excited by the incident wave (1) alone is [cf. (5) with $\theta_1 = 0$]

$$E_z^s = -\frac{2e^{1/4 i\pi}}{\sqrt{\pi}} F[\sqrt{(kr_1)}]$$

$$\sim -\frac{e^{-1/4 i\pi}}{\sqrt{\pi}} \frac{e^{-ikr_1}}{\sqrt{(kr_1)}} \left[1 + \frac{i}{2kr_1} - \frac{3}{(2kr_1)^2} + \cdots\right]$$

$$\text{for } kr \gg 1. \quad (14)$$

To begin with, the first term only of (14) is taken into account. Then as far as the excitation of $P-$ is concerned the field scattered by $P+$ can be replaced by

$$E_z = \frac{i}{\sqrt 2} H_o^{(2)}(kr_1), \qquad (15)$$

since the leading term in the asymptotic expansion of (15) is identical with that of (14).

Now the solution to the problem of a half-plane on which a cylindrical wave like (15) is incident is also well known and relatively simple (Carslaw,[8] Macdonald,[9] Clemmow[10]). Indeed, in respect of the leading term in the asymptotic development of the diffraction field, which is all that is required for present purposes, the cylindrical wave can be treated as a plane wave traveling in the direction of the line joining the cylindrical source to the edge of the half-plane and having there the same amplitude and phase.

The radiation part of the scattered field arising from the excitation of $P-$ by (15) is therefore obtained from the solution for the grazingly incident plane wave

$$E_z{}^i = -\frac{e^{-1/4i\pi}}{\sqrt \pi} \frac{e^{-ikd}}{\sqrt{(kd)}} e^{ikx} . \qquad (16)$$

In the forward direction along the negative y-axis it is

$$E_z{}^s \sim -\frac{i}{\pi} \frac{e^{-ikd}}{\sqrt{(kd)}} \frac{e^{-ikr}}{\sqrt{(kr)}} , \qquad (17)$$

and in $y = 0, x > -\frac12 d$ it is

$$E_z{}^s \sim -\frac{i}{\sqrt{2\pi}} \frac{e^{-ikd}}{(kd)} \frac{e^{-ikr_2}}{\sqrt{(kr_2)}} . \qquad (18)$$

The radiation part of the total field in the forward direction to the present degree of approximation follows by the superposition of (11) and twice (17). It is

$$E_z \sim \frac{e^{1/4i\pi}}{\sqrt{(2\pi)}} kd \left[1 + \frac{i}{kd} - 2\sqrt{\left(\frac{2}{\pi}\right)} e^{1/4i\pi} \frac{e^{-ikd}}{(kd)^{3/2}} \right] \frac{e^{-ikr}}{\sqrt{(kr)}} , \qquad (19)$$

which gives
transmission cross section =

$$d \left[1 - 2\sqrt{\left(\frac{2}{\pi}\right)} \frac{\cos(kd - \frac14\pi)}{(kd)^{3/2}} \right] . \qquad (20)$$

The formula (20) contains the first correction to the geometrical optics result (13). With very little further analysis the next correction can be similarly derived, for it is only necessary to consider the excitation of $P+$ by a cylindrical wave, diverging from the edge of $P-$, of which (18) is the radiation field. Using the same technique as that just described the radiation part of the corresponding field scattered by $P+$ in the forward direction is

[8] H. S. Carslaw, *Proc. Lond. Math. Soc.* vol. 30, p. 121 ; 1899 vol. 18, p. 291 ; 1919.
[9] H. M. Mcdonald, *Proc. Lond. Math. Soc.*, vol. 14, p. 410 ; 1915.
[10] P. C. Clemmow, *Quart. Jour. Mech. Appl. Math.*, vol. 3, p. 377 ; 1950.

$$E_z \sim \frac{e^{1/4i\pi}}{\pi\sqrt{(2\pi)}} \frac{e^{-2ikd}}{kd} \frac{e^{-ikr}}{\sqrt{(kr)}} . \qquad (21)$$

Twice (21) is therefore the next correction to (19), so that the total radiation field in the forward direction now appears as

$$E_z \sim \frac{e^{1/4i\pi}}{\sqrt{(2\pi)}} kd \left[1 + \frac{i}{kd} - 2\sqrt{\left(\frac{2}{\pi}\right)} e^{1/4i\pi} \frac{e^{-ikd}}{(kd)^{3/2}} + \frac{2}{\pi} \frac{e^{-2ikd}}{(kd)^2} \right] \frac{e^{-ikr}}{\sqrt{(kr)}} . \qquad (22)$$

Corresponding to (22),
transmission cross section

$$= d \left[1 - 2\sqrt{\left(\frac{2}{\pi}\right)} \frac{\cos(kd - \frac14\pi)}{(kd)^{3/2}} + \frac{2}{\pi} \frac{\cos(2kd)}{(kd)^2} \right]. \qquad (23)$$

The formula (23) is identical with that given by Burger[11] using a different method. It has also been previously obtained by yet another method by Dr. H. Levine, to whom the author is indebted for the communication of unpublished results.

In working out the next approximation to (23) it is evident that account must be taken in the asymptotic expansion (14) of the second term, which has not so far come into the calculation. This complicates matters somewhat, but further results have been obtained. These are stated, together with a very brief outline of the procedure, in the next section.

2.5. *Further Approximations for the Case of E-Polarization*

To the order of approximation so far considered it was adequate to replace (14) by (15). To get further approximations further cylindrical waves must be added to (15) to take account of terms in (14) other than the first: moreover, in the plane $y = 0$, the leading terms of the asymptotic expansions of these cylindrical waves must be inversely proportional to successively increasing powers of kr_1. This process must likewise be applied to the other partial scattered fields which arise at each subsequent stage of the calculation.

A series of waves of the type required can in fact be obtained by repeated differentiation with respect to ky_o of the wave $E_z = H_o^{(2)}(kR)$, where R denotes distance from $(\frac12 d, y_o)$, y_o subsequently being put equal to zero. If the nth wave of this series is denoted by K_n, (15), for example, is replaced by

$$E_z = -\frac{e^{-1/4i\pi}}{\sqrt \pi} \left(K_o + \frac{3i}{8} K_1 + \frac{105}{128} K_2 + \cdots \right), \qquad (24)$$

where, on $\theta_1 = 0$,

$$K_n(r_1, \theta_1) \sim \frac{e^{-ikr_1}}{(kr_1)^{n+1/2}}. \qquad (25)$$

Since a K_n wave is obtained by a number of differentiations of an $H_o^{(2)}$ wave, the solution for a half-plane irradiated by an K_n wave can be found by a number of differentiations of the solution for a half-plane irradi-

[11] A. P. Burger, Aeronat. Res. Inst., Rep. F157, Amsterdam; 1954.

ated by an $H_o^{(2)}$ wave, references to which were given in Section 2.4. In principle, therefore, an indefinite number of terms of the asymptotic development of the field in inverse powers of kd can be worked out. The analysis becomes progressively heavier at each stage, but the following expression for the field in the forward direction has been obtained,[12]

$$E_z \sim \frac{e^{1/4i\pi}}{\sqrt{(2\pi)}} kd \left[1 + \frac{i}{kd} - 2\sqrt{\left(\frac{2}{\pi}\right)} e^{1/4i\pi} \frac{e^{-ikd}}{(kd)^{3/2}} \right.$$

$$+ \frac{2}{\pi} \frac{e^{-2ikd}}{(kd)^2} - \frac{1}{\pi}\sqrt{\left(\frac{2}{\pi}\right)} e^{-1/4i\pi} \frac{e^{-3ikd} - \frac{7}{4}\pi e^{-ikd}}{(kd)^{5/2}}$$

$$\left. - \frac{i}{\pi^2} \frac{e^{-4ikd} - \frac{5}{2}\pi e^{-2ikd}}{(kd)^3} + O\{(kd)^{-\frac{7}{2}}\} \right] \frac{e^{-ikr}}{\sqrt{(kr)}}, \qquad (26)$$

and, correspondingly,
transmission cross section

$$= d \left[1 - 2\sqrt{\left(\frac{2}{\pi}\right)} \frac{\cos(kd - \frac{1}{4}\pi)}{(kd)^{3/2}} + \frac{2}{\pi} \frac{\cos(2kd)}{(kd)^2} \right.$$

$$- \frac{1}{\pi}\sqrt{\left(\frac{2}{\pi}\right)} \frac{\cos(3kd + \frac{1}{4}\pi) - \frac{7}{4}\pi \cos(kd + \frac{1}{4}\pi)}{(kd)^{5/2}}$$

$$\left. - \frac{1}{\pi^2} \frac{\sin(4kd) - \frac{5}{2}\pi \sin(2kd)}{(kd)^3} + O\{(kd)^{-7/2}\} \right]. \qquad (27)$$

2.6. *The Case of H-Polarization*

If we consider the correction to the analysis given in Section 2.3 for the case of H-polarization, it is at once evident that the leading correction term is a higher order in $(kd)^{-1}$ than that for E-polarization. For the field in $y = 0$, $x < \frac{1}{2}d$ scattered by $P+$ when excited by the incident wave (2) alone is

$$E_z = - Z \frac{e^{1/4i\pi}}{\sqrt{\pi}} \left\{ 2 F[\sqrt{(kr_1)}] + i \frac{e^{-ikr1}}{\sqrt{(kr_1)}} \right\}$$

$$\sim - Z \frac{e^{1/4i\pi}}{2\sqrt{\pi}} \frac{e^{-ikr1}}{(kr_1)^{3/2}} \qquad (28)$$

(where Z is the intrinsic impedance of free-space), H_z being, of course, zero; and a comparison of (28) with (14) shows that the contribution to the field arising from the interaction of $P+$ and $P-$ is less for H-polarization than for E-polarization by a factor of the order $(kd)^{-1}$.

The excitation of $P-$ by the field (28) can be worked out from a single differentiation of the known solution for the incidence on a half-plane of the cylindrical wave for which H_z is given by the $H_o^{(2)}$ function. In this way the radiation field in the forward direction containing the next approximation to (12) is found to be

$$H_z \sim \frac{e^{1/4i\pi}}{\sqrt{(2\pi)}} kd \left[1 - \frac{i}{kd} + \frac{e^{-1/4i\pi}}{\sqrt{(2\pi)}} \frac{e^{-ikd}}{(kd)^{5/2}} \right.$$

$$\left. + O\{(kd)^{-7/2}\} \right] \frac{e^{-ikr}}{\sqrt{(kr)}}, \qquad (29)$$

associated with
transmission cross section

$$= d \left[1 + \frac{1}{\sqrt{(2\pi)}} \frac{\cos(kd + \frac{1}{4}\pi)}{(kd)^{5/2}} + O\{(kd)^{-\frac{7}{2}}\} \right]. \qquad (30)$$

The formula (30) has also been previously obtained by Dr. H. Levine (unpublished) using an independent method.

2.7 *Numerical Results*

In this section it is convenient to talk about the transmission *coefficient* rather than the transmission *cross section*. The former is just the latter divided by d.

The formulas (23) and (30), divided by d, should give good approximations to the transmission coefficients for E-polarization and H-polarization respectively for $kd \gg 1$. Evidently as kd increases indefinitely they both tend to unity in an oscillatory manner, but the latter approaches this limit much the more rapidly of the two.

Using an eigenfunction method, Skavlem[13] has been able to make calculations of the transmission coefficients which are apparently exact to five figures for values of kd up to 20. For E-polarization, Burger[11] has shown that the numerical values yielded by (23) agree well with those given by Skavlem, the discrepancy only becoming greater than about 1 per cent for $kd < 4$.

Likewise, calculations based on (30) agree well with the relatively few given by Skavlem for H-polarization, the discrepancy only becoming greater than about 1 per cent for $kd < 3$. In this case, however, Skavlem appears to have been misled by having made calculations with $kd > 3$ only for the values 4, 8, 16; for he remarks ".... the transmission coefficient varies in a rather monotonic way with only one minimum and no maxima," and illustrates with a figure showing the variation going in this way. In fact the earlier results of Morse and Rubenstein,[14] although confined to values of kd less than about 8, indicate that the transmission coefficient does oscillate about unity in accordance with (30).

3. The Circular Aperture

Quite recently a number of measurements of the field in the plane and on the axis of a circular aperture on which a plane wave is normally incident have been reported, and various methods have been suggested for predicting these fields. In particular, Andrejewski[5] has made computations, based on the rigorous solution in terms of spheroidal functions, for some values of ka up to 10, where a is the radius of the aperture.

The method discussed in the present paper, in its crudest form neglecting interaction, has been applied to this problem by Mr. R. F. Millar. The details of his analysis, together with numerical calculations on apertures of different radii, will be reported independently

---

[12] The author does not regard it as inconceivable that there should be an error in the last two terms of (26) arising from a slip in the calculation.

[13] S. Skavlem, *Arch. Math. Naturvid.*, 51, p. 61 ; 1951.
[14] P. M. Morse and P. J. Rubenstein, *Phys. Rev.*, vol. 54, p. 895 ; 1938.
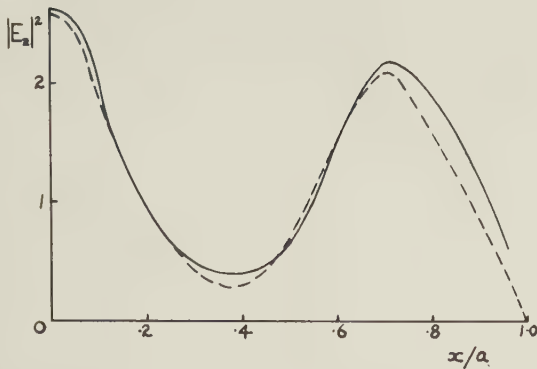
Fig. 2 — $| E_z |^2$ as a function of x/a for $ka = 10$, $y = z = 0$. Solid curve = present theory, dotted curve = exact calculation by Andrejewski.

Fig. 3 — $| E_z |^2$ as a function of $-y/\lambda$ for $ka = 3\pi$, $x = z = 0$

Solid curve = present theory, solid dots = experimental points due to Andrews, circles = experimental points due to Ehrlich, Silver, and Held.

elsewhere. But as an indication of the extent to which the method succeeds typical results are shown in Figs. 2 and 3.

Let the plane of the screen be $y = 0$, the aperture $x^2 + z^2 = a^2$, and the incident plane wave $\boldsymbol{E}^i = (0, 0, e^{iky})$. Then in figure 2 $| E_z |^2$ in $y = 0$ is plotted against x/a, for $z = 0$, when $ka = 10$. The full line curve is obtained from the present theory by a numerical evaluation of the integrals involved, whilst the dashed curve is that due to Andrejewski[5] and may be reckoned exact.

In Fig. 3 $| E_z |^2$ on the negative y-axis is plotted against $-y/\lambda$ for $ka = 3\pi$ (λ being the wavelength). The full line curve is obtained from the relatively simple formula given by the present theory, the solid dots are from measurements by Andrews[4] and the open dots from measurements by Ehrlich, Silver and Held.[15]

It has already been remarked that the formulas for the field in the aperture given by the present method can be compared directly with those suggested by Andrews.[4] They differ somewhat. And the point where the effect of the difference seems to be greatest is at the center of the aperture. Here the present method gives (in agreement with Frahn[3])

$$E_z = 1 - \frac{1}{\sqrt{2}} e^{-ika}, \tag{31}$$

whilst Andrews' formula, and also Bekefi's[16] method, give

$$E_z = 1 - \tfrac{1}{2} e^{-ika}. \tag{32}$$

It must be admitted that by and large the experimental results (for example, those of Andrews[4] and of Ehrlich, Silver and Held[15]) seem to favor (32) rather than (31). The latter, however, is closer to Andrejewski's exact value for $ka = 10$, and should, one suspects, be the limit of the correct expression as $ka \rightarrow \infty$.

4. COMMENTS

It is apparent that some more general results than those mentioned here could be obtained. For instance, in the case of the slit the whole analysis could be carried through for an obliquely incident plane wave, or for line or point sources. Again, for the circular aperture a better estimation of the edge currents might be derived by treating each pair of diametrically opposite edges as a slit, thus allowing to some extent for interaction across the aperture. Unlike the approximate method discussed in Section 3, this should lead to a correction term to the transmission cross section of geometrical optics. Clearly, also, other smooth apertures, such as an elliptic one, could be treated.

5. ACKNOWLEDGMENT

[15] M. J. Ehrlich, S. Silver, and G. Held, *Jour. Appl. Phys.*, vol. 26, p. 336; 1955.
[16] G. Bekefi, *Jour. Appl. Phys.*, vol. 24, p. 1123; 1953.

On Discontinuous Electromagnetic Waves and the Occurrence of a Surface Wave

B. VAN DER POL†

Summary—Two problems are considered: (1) The field around a dipole free in space. Contrary to the usual treatment, where the moment of the dipole is considered to vary harmonically in time, here the moment is assumed initially to be zero but at the instant $t = 0$ to jump to a constant value, which it further maintains. (2) The same dipole is placed vertically on a horizontal plane separating two media of different refractive index. It is shown that the resulting disturbance on the plane is composed of two space waves and one surface wave. First the Hertzian vector at a distance ρ from the dipole is zero. At $t = t_1$ the disturbance arrives there through the less dense medium, and slowly begins to rise till, at the moment $t = t_2$, when the disturbance has had time to reach the same distance through the second (denser) medium it reaches its final static value and further stays constant. During the transitory interval $t_1 < t < t_2$ the disturbance is found to be representable, apart from a constant, by a pure surface wave.

The two problems are solved with the help of the modern form of the operational calculus based on the two-sided Laplace transform. The analytical tools of the operational calculus needed are explained in a separate paragraph.

INTRODUCTION

I**N 1888, Hertz, in a classical paper,[1] gave a solution of Maxwell's equation for free space, with a singularity (simple pole) at the origin. This solution is at the base of all modern calculations concerning the radiation distribution of antennas.

In his derivation Hertz made use of rectangular co-ordinates. A more modern version of the same problem, but in spherical co-ordinates, was published by Sommerfeld[2] in 1948.

The problem in question, in modern terms, concerns the field round an infinitesimal dipole of moment $P = e \cdot l$, formed by two equal and opposite charges $+e$ and $-e$ at a distance l apart.

Lorentz[3] encountered the same mathematical problem in his theory of electrons.

It is usual in this problem to assume the moment P of the infinitesimal dipole to vary harmonically with time, thus simulating the oscillatory moment of a radio antenna.

It is the purpose of the present note to study two problems:

1. The field in free space round a dipole whose moment P, assumed initially to be zero, but which, at the time $t = 0$, suddenly jumps to a constant value 1, which it further maintains. This dipole moment therefore has the mathematical representation $P = U(t)$ where $U(t)$ is the Heaviside unit function as represented in Fig. 1.

† International Telecommun. Union, Geneva, Switzerland.
[1] H. Hertz, *Ann. der Phys.* vol. 36, p. 1; 1888—Gesammelte Werke II, vol. 147, Leipzig, Ger.; 1914.
[2] A. Sommerfeld, "Vorlesungen ueber theoretische physik, band III," *Elektrodynamik*, pp. 150–155; Wiesbaden, Ger.; 1948.
[3] H. A. Lorentz, "The Theory of Electrons," 2nd ed., p. 56; Leipzig, Ger.; 1916.
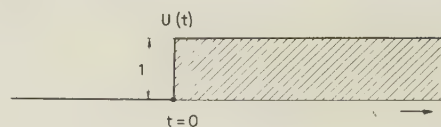
Fig. 1

Physically this may be. interpreted as the sudden separation, to a certain small distance, of two opposite charges which originally coincided; and the study of the resulting electric and magnetic fields. The very different behaviour of these fields near the dipole and further away thus becomes more evident than from the usual oscillatory treatment.

2. The problem in which this same dipole, with a discontinuous moment, is placed vertically on an infinite horizontal plane, which separates two media of different refractive index.

The latter problem is analogous to, say, a vertical transmitter placed on a flat earth, with the difference that the dipole moment of the transmitter does not vary harmonically with time, but is represented by $U(t)$.

The solution of the latter problem shows the presence in the plane $z = 0$ of a typical surface wave, *i.e.*, a wave satisfying the two-dimensional wave equation.

Contrary to the usual treatment of an oscillatory dipole, where extensive use is made of complex function theory, the present problems (1) and (2) both appear to be amenable to a treatment in the real domain only, in the sense that no complex integrations are necessary.

In the following we will make a free use of the ideas and methods of the modern form of the operational calculus,[4] as developed in the book quoted below.

A short resumé of the operational tools needed here is given in the next paragraph.

OPERATIONAL METHODS

If a function of the time $h(t)$ be given

$$(-\infty < t < +\infty)$$

and which we call the "original," we obtain its "image" $f(p)$ by means of the two-sided Laplace transform

$$f(p) = p \int_{-\infty}^{+\infty} e^{-pt} h(t) \, dt, \tag{1}$$

provided that the function $h(t)$ is such that there exists in the complex p-plane a band $\alpha < \mathrm{Re} p < \beta$ in which this integral converges.

[4] B. van der Pol and H. Bremmer, "Operational Calculus Based on the Two-Sided Laplace Transform," Cambridge Univ. Press, London, Eng.; 1950.

We write the relation (1) in short

$$f(p) \doteqdot h(t), \qquad \alpha < \mathrm{Re}\,p < \beta,$$

or

$$h(t) \doteqdot f(p), \qquad \alpha < \mathrm{Re}\,p < \beta.$$

If $h(t)$ happens to be different from zero only on the interval $a < t < b$ then the integral (1) áutomatically reduces to

$$f(p) = p \int_a^b e^{-pt}\, h(t)\, dt.$$

In particular, if $h(t) = U(t)$, then

$$f(p) = p \int_{-\infty}^{+\infty} e^{-pt}\, U(t)\, dt = p \int_0^{+\infty} e^{-pt}\, 1\, dt = 1, \quad (1a)$$

so that we have the simple correspondence:

$$1 \doteqdot U(t), \qquad \mathrm{Re}\,p > 0, \tag{2}$$

or, in words, the unit function $U(t)$ has, as its image in the p field, the function 1.

From (1) it follows that, if

$$f(p) \doteqdot h(t), \qquad \alpha < \mathrm{Re}\,p < \beta,$$

then the following "shift rule" is generally applicable

$$e^{-\lambda p} f(p) \doteqdot h(t - \lambda), \qquad \alpha < \mathrm{Re}\,p < \beta, \tag{3}$$

where λ is a positive or negative real constant. Incidentally, this is not generally true if the operational calculus is based on a one-sided Laplace transform.

From (3) it follows at once that the retarded (or advanced) unit function has the following image

$$e^{-\lambda p} \doteqdot U(t - \lambda), \qquad \mathrm{Re}\,p > 0.$$

Similarly the Dirac delta function $\delta(t)$ has as its image

$$f(p) = p \int_{-\infty}^{+\infty} e^{-pt} \delta(t)\, dt = p \int_{-\infty}^{+\infty} 1\, \delta(t)\, dt = p,$$

and hence

$$p \doteqdot \delta(t), \qquad -\infty < \mathrm{Re}\,p < +\infty.$$

A Dipole of Moment $P = U(t)$, in Free Space

We first solve this, rather simple, introductory problem, as the ideas used in its solution will be applied later on to the more complicated problem discussed in the next section.

The basic formulas for the magnetic force H_φ and the electric components E_r and E_θ around a dipole of a general moment P (in the usual notations) are to be found in Sommerfeld's lectures.[2] They can be written in the following form:

$$\begin{cases} \dfrac{4\pi}{c} H_\varphi = \sin\theta \left\{ 0 \quad + \dfrac{1}{r^2}\dfrac{\partial P}{\partial(ct)} + \dfrac{1}{r}\dfrac{\partial^2 P}{\partial(ct)^2} \right\} \\[2ex] 4\pi E_r = 2\cos\theta \left\{ \dfrac{1}{r^3} P + \dfrac{1}{r^2}\dfrac{\partial P}{\partial(ct)} + 0 \right\} \\[2ex] 4\pi E_\theta = \sin\theta \left\{ \dfrac{1}{r^3} P + \dfrac{1}{r^2}\dfrac{\partial P}{\partial(ct)} + \dfrac{1}{r}\dfrac{\partial^2 P}{\partial(ct)^2} \right\} . \end{cases} \tag{4}$$

The argument of P has to be taken retarded as $P = P(t - r/c)$. These field components are derived there by differentiation from a Hertzian vector Π (which derivation will not be reproduced here). This vector is given by

$$4\pi\Pi = \frac{1}{r} P\left(t - \frac{r}{c}\right). \tag{5}$$

It satisfies the wave equation

$$\Delta\Pi = \frac{1}{c^2}\frac{\partial^2 \Pi}{\partial t^2}, \qquad (r \neq o). \tag{6}$$

If we introduce, instead of t, the variable τ, given by

$$ct = \tau,$$

and if we transform τ operationally, so that, e.g.,

$$\tau U(\tau) \doteqdot \frac{1}{p} \qquad \mathrm{Re}\,p > 0, \tag{6a}$$

and if we assume Π to have the image $f(p)$, then (6) becomes

$$(\Delta - p^2)f(p) = 0 \qquad (r \neq 0). \tag{7}$$

If the moment P is a unit function, (5) becomes

$$4\pi\Pi = \frac{1}{r} P\left(t - \frac{r}{c}\right) = \frac{1}{r} U\left(\frac{\tau - r}{c}\right) = \frac{1}{r} U(\tau - r) \doteqdot \frac{1}{r} e^{-pr}.$$

Further (7) shows that we immediately obtain the operational image for our functions (such as $\Pi(x, y, z, t)$ or the field components of E and H) from the usual expressions, pertaining to an oscillating dipole with time factor $e^{-i\omega t}$, if with the standard notations, we replace in (4) $\partial/\partial t$ by $-i\omega$, then write $-i\omega/c = -ik$ and finally replace ik by $-p$. This procedure amounts to the same as replacing directly $\partial/(\partial(ct))$ by p.

Thus from (4) we obtain the following images.

$$\begin{cases} \dfrac{4\pi}{c} H_\varphi \doteqdot e^{-pr} \sin\theta \left\{ 0 + \dfrac{p}{r^2} + \dfrac{p^2}{r} \right\} \\[2ex] 4\pi E_r \doteqdot e^{-pr} 2\cos\theta \left\{ \dfrac{1}{r^3} + \dfrac{p}{r^2} + 0 \right\} \\[2ex] 4\pi E_\theta \doteqdot e^{-pr} \sin\theta \left\{ \dfrac{1}{r^3} + \dfrac{p}{r^2} + \dfrac{p^2}{r} \right\}, \end{cases} \tag{4a}$$

where the term e^{-pr} takes account of the necessary retardation.

Hence (4a) constitutes the final solution of problem (1) in operational form. It remains to be interpreted in the τ (or t) field. With the aid of the exposition in the previous section we find at once the "translation" of (4a) to be

$$\begin{cases} \dfrac{4\pi}{c} H_\varphi = \sin\theta \left\{ 0 \qquad\qquad + \dfrac{1}{r^2}\delta(\tau - r) \right. \\[2ex] 4\pi E_r = 2\cos\theta \left\{ \dfrac{1}{r^3}U(\tau - r) + \dfrac{1}{r^2}\delta(\tau - r) \right. \\[2ex] 4\pi E_\theta = \sin\theta \left\{ \dfrac{1}{r^3}U(\tau - r) + \dfrac{1}{r^2}\delta(\tau - r) \right. \end{cases}$$

$$+ \frac{1}{r}\delta'(\tau - r) \Big\}$$

$$+ 0 \qquad\qquad \Big\} \qquad (8)$$

$$+ \frac{1}{r}\delta'(\tau - r) \Big\}.$$

Thus we see that, for a moment $P = U(t)$, and near the dipole (and therefore considering the terms with r^{-3} only), both E_r and E_θ contain a term of the form $U(t)$, i.e. a unit function similar to the time function of the moment of the dipole itself. This corresponds to the static solution. But all three components H_φ, E_r and E_θ also contain a term with a delta function $\delta(\tau - r)$. For large distances (which correspond to what is usually called the radiation field) only the terms containing a factor r^{-1} are of importance. They are present in the components E_θ and H_φ only, and contain as time functions the "derivative" of a delta function $\delta'(\tau - r)$.

These results are shown in Fig. 2, where the two (or three) time components present in the different field components are represented symbolically. As we saw above, their relative importance in each case depends on the distance r from the transmitter.

It therefore follows that what we notice at a large distance from the dipole (the terms with factor r^{-1} only) corresponds to the *second* derivative of the electrical moment of the dipole.

This fact is thus much more clearly brought out for a dipole moment of the form $U(t)$ than for a sinusoidal excitation.

A Dipole of Moment $U(t)$ Placed Vertically on a Horizontal Plane which Separates Two Different Media.

This problem is classical in radio when the moment of the dipole is considered to vary harmonically in time. It was first solved by Sommerfeld[5] in 1909. However here we wish to treat the case where the moment is a unit function $U(t)$. Again, we won't give the full derivation of the problem here, but shall borrow from Sommerfeld's investigation (as expounded, e.g., in his lectures on electromagnetic theory[2]) the solution for an oscillating dipole.

Further it will appear that all the essential points of our solution are already brought out if we do not introduce any absorption either in the first or the second

$M = e\ell$:   U(t)

H_φ: $\delta(t)$ $\delta'(t)$

E_r: U(t) $\delta(t)$

E_θ: U(t) $\delta(t)$ $\delta'(t)$
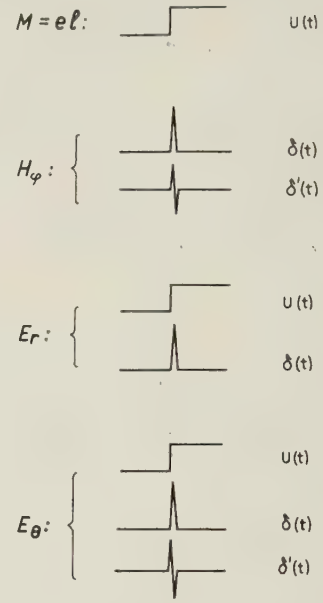
Fig. 2

medium, thus considerably simplifying our analysis. The present problem therefore may be visualized as the propagation of waves from a discontinuous point source, placed at the plane separating, e.g., an upper half space which is empty and a lower half space of glass or a similar substance.

With Sommerfeld, we introduce again a Hertzian vector having a vertical component only. In the first (upper) medium with a refractive index $n_1^2 = \epsilon_1$, it is

$$\Pi = \Pi_1$$

and solves the wave equation pertaining to the first medium (with k_1^2)

$$(\Delta + k_1^2)\Pi_1 = 0.$$

Similarly, in the second (lower) medium it equals

$$\Pi = \Pi_2$$

and there it solves

$$(\Delta + k_2^2)\Pi_2 = 0.$$

The boundary conditions, together with the condition at infinity yield as the classical solution in the first medium $(z > 0)$ in cylindrical co-ordinates and with the classical notations $\rho^2 = x^2 + y^2$, $r^2 = \rho^2 + z^2$:

$$\frac{\Pi_1(\rho, z)}{k_1^2 + k_2^2} = \int_o^\infty e^{-z\sqrt{\lambda^2 - k_1^2}}\, \frac{J_o(\lambda\rho)}{N}\lambda d\lambda. \qquad [6] \qquad (9)$$

[6] For reasons of symmetry we have here changed the normalization slightly. Sommerfeld, in his lectures, uses the factor $1/2k_2^3$ where above, in the first member of (9), we take $(k_1^2 + k_2^2)^{-1}$.

[7] B. van der Pol, *Jahr. fur Draht. Tel. u Tel.* vol. 37, pp. 152–156; 1931.

[5] A. Sommerfeld, *Ann. der Phys.*, vol. 28, p. 287; 1909—and vol. 81, p. 1135; 1926.

In (9) the denominator N is given by

$$\frac{1}{N} = \frac{1}{k_2{}^2 \sqrt{\lambda^2 - k_1{}^2} + k_1{}^2 \sqrt{\lambda^2 - k_2{}^2}}.$$

Throughout the square roots are defined as having their real part positive.

Before proceeding to use operational methods for the discontinuous excitation, we first introduce in (9) an artifice we found in 1931[7] and which will appear to be basic for the following treatment. It consists of writing $1/N$ as the following integral (which is easily verified)

$$\frac{1}{N} = \frac{k_1 k_2}{k_1{}^4 - k_2{}^4} \int_{u=k_1}^{u=k_2} \frac{1}{\sqrt{\lambda^2 - u^2}} \, d\left(\frac{1}{\sqrt{\dfrac{u^2}{h^2} - 1}}\right), \quad (10)$$

where

$$\frac{1}{h^2} = \frac{1}{k_1{}^2} + \frac{1}{k_2{}^2}. \quad (10a)$$

Hence the complete solution in the first medium can be given the form of a double integral:

$$\Pi_1(\rho, z) = \int_{\lambda=0}^{\lambda=\infty} e^{-z\sqrt{\lambda^2-k_1{}^2}} \, J_o(\lambda\rho) \left\{ \frac{k_1 k_2}{k_1{}^2 - k_2{}^2} \right.$$
$$\left. \int_{u=k_1}^{u=k_2} \frac{1}{\sqrt{\lambda^2 - u^2}} \, d\left(\frac{1}{\sqrt{\dfrac{u^2}{h^2} - 1}}\right)\right\} \lambda d\lambda. \quad (11)$$

This is the form of the solution for an oscillating dipole. We now transform it into the operational image pertaining to a dipole of moment $U(t)$. Again, as in problem (1) above, we introduce the variable $\tau = ct$. We again transform τ operationally as in (6a) and assume $\mathrm{Re}\,p > 0$. We further introduce the (real) refractive indices $n_1{}^2 = \epsilon_1$, $n_2{}^2 = \epsilon_2$. As before, we therefore have to replace

$$ik_1 \quad \text{by} \quad -pn_1,$$
$$ik_2 \quad \text{by} \quad -pn_2,$$

and, according to (10a),

$$ih \quad \text{by} \quad -p\frac{n_1 n_2}{\sqrt{n_1{}^2 + n_2{}^2}}.$$

The latter refractive index, corresponding to h, we shall further write

$$\frac{n_1 n_2}{\sqrt{n_1{}^2 + n_2{}^2}} = n_{12}. \quad (11a)$$

Applying these transformations to (11) and changing at the same time the integration variable u to $u = ip\nu$, we obtain the following image for $\Pi_1(\rho, z)$: $(\mathrm{Re}\,p > 0)$

$$\Pi_1(\rho, z) \doteqdot \int_{\lambda=0}^{\infty} e^{-z\sqrt{\lambda^2 + p^2 n_1{}^2}} \, J_o(\lambda\rho) \left\{ \frac{n_1 n_2}{n_1{}^2 - n_2{}^2} \right.$$
$$\left. \int_{\nu=n_1}^{\nu=n_2} \frac{1}{\sqrt{\lambda^2 + p^2\nu^2}} \, d\left(\frac{1}{\sqrt{\dfrac{\nu^2}{n_{12}{}^2} - 1}}\right)\right\} \lambda \, d\lambda. \quad (12)$$

Thus (12) constitutes the complete solution, in operational form, of our problem (2), and it remains to be interpreted in the original language for $\tau = ct$.

The latter problem is materially simplified, while still maintaining all essential points, if we limit ourselves to a consideration of (12) for z approaching zero from above: $z = +0$, i.e. of the behaviour of $\Pi_1(\rho, 0)$ in the plane separating the two media. We thus find from (12)

$$\Pi_1(\rho, 0) \doteqdot \frac{n_1 n_2}{n_1{}^2 - n_2{}^2} \int_{\nu=n_1}^{\nu=n_2} d\left(\frac{1}{\sqrt{\dfrac{\nu^2}{n_{12}{}^2} - 1}}\right)$$
$$\int_{\lambda=0}^{\infty} \frac{J_o(\lambda\rho)}{\sqrt{\lambda^2 + p^2\nu^2}} \, \lambda \, d\lambda. \quad (13)$$

The latter integral is known and equals $(1/\rho)e^{-p\nu\rho}$, so that (13) becomes

$$\Pi_1(\rho, 0) \doteqdot \frac{n_1 n_2}{n_1{}^2 - n_2{}^2} \int_{\nu=n_1}^{\nu=n_2} \frac{1}{\rho} e^{-p\nu\rho} \, d\left(\frac{1}{\sqrt{\dfrac{\nu^2}{n_{12}{}^2} - 1}}\right). \quad (14)$$

We now integrate (14) by parts, using (11a), and obtain

$$\Pi_1(\rho, 0) \doteqdot \frac{n_1 n_2}{n_1{}^2 - n_2{}^2} \left\{ \frac{e^{-p\nu\rho}}{\rho} \frac{1}{\sqrt{\dfrac{\nu^2}{n_{12}{}^2} - 1}} \Bigg|_{\nu=n_1}^{\nu=n_1} \right.$$
$$\left. + p \int_{\nu=n_1}^{\nu=n_2} \frac{e^{-p\nu\rho}}{\sqrt{\dfrac{\nu^2}{n_{12}{}^2} - 1}} \, d\nu \right\} = \frac{n_1 n_2}{n_1{}^2 - n_2{}^2} \left\{ \frac{n_1}{n_2} \frac{e^{-pn_2\rho}}{\rho} \right.$$
$$\left. - \frac{n_2}{n_1} \frac{e^{-pn_1\rho}}{\rho} + p \int_{x=n_1\rho}^{x=n_2\rho} \frac{e^{-px}}{\sqrt{\dfrac{x^2}{n_{12}{}^2} - \rho^2}} \, dx \right\}, \quad (15)$$

where, in the last integral, we have replaced the integration variable ν by $x = \nu\rho$.

With the aid of the operational data contained in the second section above, it is now an easy matter to find the "original" of (15). We further assume that the optical density of the second medium is larger than that of the first, i.e., $n_2 > n_1$. From (3a) we see that the first term in the bracket of (15) has as its original

$$\frac{n_1}{n_2\rho}\, U(\tau - n_2\rho),$$

which function, therefore, is zero for $\tau < n_2\rho$, but at $\tau = n_2\rho$, it jumps to the (subsequently constant) value $(n_1/n_2)\, 1/\rho$. Similarly for the second term in (15). As to the last term in (15), this happens to have exactly the form of our definition integral as given by (1) and (1a). As here p relates to τ, the original has to be interpreted as the function which equals

$$\frac{1}{\sqrt{\dfrac{\tau^2}{n_{12}{}^2} - \rho^2}}$$

as long as τ is comprised between the limits $\tau = n_1\rho$ and $\tau = n_2\rho$, whereas the function equals zero outside this interval. With the aid of our $U(t)$ function, we can therefore write for the original of the last term of (15)

$$\frac{1}{\sqrt{\dfrac{\tau^2}{n_{12}{}^2} - \rho^2}} \cdot U(\tau - n_1\rho) \cdot U(n_2\rho - \tau). \tag{16}$$

Hence the complete original of (15) becomes

$$\Pi_1(\rho, 0) = \frac{n_1 n_2}{n_2{}^2 - n_1{}^2} \left\{ \frac{n_2}{n_1}\, \frac{1}{\rho}\, U(\tau - n_1\rho) \right.$$

$$- \frac{n_1}{n_2}\, \frac{1}{\rho}\, U\ (\tau - n_2\rho)$$

$$\left. - \frac{1}{\sqrt{\dfrac{\tau^2}{n_{12}{}^2} - \rho^2}}\, U(\tau - n_1\rho) \cdot U(n_2\rho - \tau) \right\}, \tag{17}$$

where, as before, $\tau = ct$.

Clearly the first two terms correspond to "space waves" which here degenerate into two discontinuous jumps, first at $\tau = n_1\rho$ upwards with a magnitude $n_2{}^2/(n_2{}^2 - n_1{}^2)\, 1/\rho$ and later, at $\tau = n_2\rho$ downwards with a magnitude $n_1{}^2/(n_2{}^2 - n_1{}^2)\, 1/\rho$. After $\tau = n_2\rho$ we are therefore left with a constant magnitude

$$\frac{n_2{}^2 - n_1{}^2}{n_2{}^2 - n_1{}^2}\, \frac{1}{\rho} = \frac{1}{\rho}.$$

These jumps of the two space waves (at a fixed distance ρ from the dipole) but as a function of $\tau = ct$ are depicted at the top of Fig. 3, where, at the instant $t_1 = n_1\rho/c$ the first space wave is seen to cause a jump upwards and at the later instant $t_2 = n_2\rho/c$ the second space wave is seen to cause a second jump, but of different magnitude, downwards.

As to the last term $(\tau^2/n_{12}{}^2 - \rho^2)^{-1/2}$ in (17), this clearly represents a typical surface wave, as it has the form of a solution of the *two*-dimensional wave equation, which, as is well known, corresponds to the three-dimensional potential equation, if in this latter we
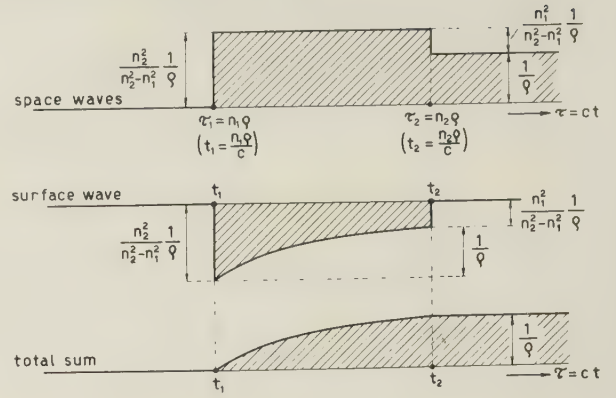
Fig. 3

change one linear dimension by ict. Thus a three-dimensional potential is

$$(x^2 + y^2 + z^2)^{-1/2} = (\rho^2 + z^2)^{-1/2},$$

and hence a two-dimensional (surface) wave is

$$(\rho^2 - c^2t^2)^{-1/2} = (\rho^2 - \tau^2)^{-1/2},$$

which is exactly of the form found here. Moreover the refractive index associated with the surface wave is given by $n_{12} = n_1 n_2/\sqrt{n_1{}^2 + n_2{}^2}$ which is symmetrically built up of the two refractive indices n_1 and n_2 of the two corresponding media. The latter fact constitutes another reason for calling the last term of (17) a surface wave. Now let us investigate its behavior. According to the factor $U(\tau - n_1\rho)$ it comes into being only at $\tau = n_1\rho$, i.e., at $t = t_1 = n_1\rho/c$.

At this instant it suddenly jumps to the value

$$(\tau^2/n_{12}{}^2 - \rho^2)^{-1/2} = \left(\frac{n_1{}^2}{n_{12}{}^2}\, \rho^2 - \rho^2\right)^{-1/2}$$

$$= \frac{1}{\rho} \left(\frac{n_1{}^2(n_1{}^2 + n_2{}^2)}{n_1{}^2 n_2{}^2} - 1\right)^{-1/2} = \frac{n_2}{n_1}\, \frac{1}{\rho},$$

or, with the over-all factor taken into account, it jumps downwards at $t_1 = n_1\rho/c$ from zero to the value

$$- \frac{n_2{}^2}{n_2{}^2 - n_1{}^2}\, \frac{1}{\rho}.$$

This jump is exactly equal in magnitude, but of opposite sense, to the jump at the same instant made upwards by the space wave. Afterwards the surface wave is for a certain duration represented by

$$- \frac{n_1 n_2}{n_2{}^2 - n_1{}^2}\, (\tau^2/n_{12}{}^2 - \rho^2)^{-1/2} \tag{18}$$

but at $\tau = n_2\rho$ corresponding to $t = t_2 = n_2\rho/c$ it must suddenly disappear, according to the last factor $U(n_2\rho - \tau)$, because, after this time, the argument of this U-function becomes negative and therefore the function itself becomes zero. At this latter instant $t = t_2$ the value of (18) has become

$$- \frac{n_1 n_2}{n_2{}^2 - n_1{}^2} \left(\frac{n_2{}^2}{n_{12}{}^2} \rho^2 - \rho^2 \right)^{-1/2}$$

$$= - \frac{n_1 n_2}{n_2{}^2 - n_1{}^2} \frac{1}{\rho} \left(\frac{n_2{}^2(n_1{}^2 + n_2{}^2)}{n_1{}^2 n_2{}^2} - 1 \right)^{-1/2}$$

$$= - \frac{n_1{}^2}{n_2{}^2 - n_1{}^2} \frac{1}{\rho},$$

and, as it has suddenly to become zero, its jump upwards equals the latter amount. But, as we saw, this amount also equals the jump downwards by the space wave at the same instant $t = t_2$. Hence the two pairs of discontinuities compensate each other.

The described behavior of the surface wave is depicted in the middle of Fig. 3. At the bottom of the same figure the total function (17) representing the addition of the two space waves and the surface wave is also shown.

From the above it is seen that the total phenomenon behaves as follows:

1. Before the moment $t = t_1$ $(\tau = n_1 \rho)$ the function $\Pi(\rho, 0)$ is strictly zero.

2. At this moment it begins to rise, without any discontinuity in the function itself, according to the expression

$$\frac{n_2{}^2}{n_2{}^2 - n_1{}^2} \left\{ \frac{1}{\rho} - \frac{n_1/n_2}{(\tau^2/n_{12}{}^2 - \rho^2)^{1/2}} \right\}, \qquad (19)$$

which, in its time behavior, is typical for a surface wave.

3. At the time $t = t_2$ $(\tau = n_2 \rho)$, when the disturbance has had time to reach the distance ρ through the second (denser) medium, (19) has become equal to $1/\rho$ this being the final steady value which is further maintained and which corresponds to the electrostatic potential theory.

A physical explanation of the phenomena may be given in the following way. At the moment $t = t_1$, when the disturbance first reaches the distance ρ through the first (less dense) medium, it cannot immediately obtain its final steady value $1/\rho$, because energy will initially be lost to the second (denser) medium. This process is only terminated when sufficient time has elapsed (up to $t = t_2$), after which the disturbance has reached the same distance ρ through the second medium. Thereupon the final steady value of $1/\rho$ is maintained.

It is of interest that in the transitory period $t_1 < t < t_2$ the behavior of $\Pi_1(\rho, 0)$ is, apart from a constant term, entirely governed by a pure surface wave.

Finally, as to the normalisation introduced above, it is easy to note from (15) that the behavior of $\Pi(\rho, 0)$ near $\rho \to 0$ is as follows:

$$\operatorname*{Lim}_{\rho \to 0} \quad \rho \cdot \Pi(\rho, 0)$$

$$\doteq \frac{n_1 n_2}{n_2{}^2 - n_1{}^2} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} + 0 \right) = 1 \doteq U(t).$$

Hence, very near $\rho = 0$, $\Pi(\rho, 0)$ behaves as

$$\Pi(\rho, 0) = \frac{1}{\rho} U(t),$$

and this expression correctly describes the unit function behavior of the dipole moment.

We shall not derive here the field components H_φ, E_r and E_θ (as was done in our first problem) as differentiations of $\Pi_1(\rho, z)$ with respect to either the time or the co-ordinates does not materially change its character of a surface wave in the transitory period.

In conclusion it may thus be said that during the transitory period $t_1 < t < t_2$ the surface wave is very much in evidence. In fact, apart from a constant, the surface wave is the only disturbance present.

The Excitation of a Perfectly Conducting Half-Plane by a Dipole Field[*]

A. E. HEINS†

Summary—Starting with the solution of two scalar problems in diffraction theory derived by MacDonald in 1915, it is shown that the following problem may be solved. An electric or magnetic dipole is situated in the presence of a semi-infinite, perfectly conducting, thin plane. This problem may be solved by appealing to an appropriate representation of the electromagnetic field. When the formulation is complete, we are left merely with a two-dimensional Poisson equation. The method serves to show why some orientations of the dipole are simpler to handle than others.

INTRODUCTION

IT HAS recently been demonstrated by Senior[1] that one may discuss the effect of an electromagnetic dipole field in the presence of a semi-infinite, perfectly conducting half plane of zero thickness. The analysis of the problem is carried through by synthesizing the corresponding plane wave solution. Surprisingly enough, this problem had not been treated adequately before Senior, since among other things, the nature of the field at the edge of the half-plane had not been taken into account. On this score, a minor modification of the recent work of Heins and Silver[2] would show that with the use of an appropriate representation, the field components would have the appropriate orders at the edge.

We shall show here that we may derive the solution to this vector problem by appealing to the solution of two known scalar problems. This can be done by using a representation of the electromagnetic field which permits the determination of the components of the electric and magnetic fields parallel to the edge of the half plane in terms of the two scalar problems which we shall describe later.[3] It will then be found that the other component of the electric field tangent to the plane is determined up to an unknown function which satisfies the Poisson equation in two variables. The method we employ, demonstrates, incidentally, why the cases in which there are no components of the electric or magnetic fields parallel to the edge of the half plane are so simple.

FORMULATION OF THE PROBLEM

Consider a perfectly conducting half-plane of zero thickness, $-\infty < x < \infty, y > 0, z = 0$ in whose presence there is located an electric dipole whose electric field components may be written as

$$\text{(a)}\quad \frac{\partial^2 F}{\partial x^2} + k^2 F, \qquad \frac{\partial^2 F}{\partial x \partial y}, \qquad \frac{\partial^2 F}{\partial x \partial z} \qquad (1a)$$

$$\text{(b)}\quad \frac{\partial^2 F}{\partial y \partial x}, \qquad \frac{\partial^2 F}{\partial y^2} + k^2 F, \qquad \frac{\partial^2 F}{\partial y \partial z} \qquad (1b)$$

$$\text{(c)}\quad \frac{\partial^2 F}{\partial z \partial x}, \qquad \frac{\partial^2 F}{\partial y \partial z}, \qquad \frac{\partial^2 F}{\partial z^2} + k^2 F \qquad (1c)$$

Here $F = \exp[ik \mid \boldsymbol{r} - \boldsymbol{r}_o \mid]/k \mid \boldsymbol{r} - \boldsymbol{r}_o \mid$ and $\boldsymbol{r}_o$ is the vector whose components are (x_o, y_o, z_o) and locates the source point. We note the absence of a component of the magnetic field in each of the cases in (1).

We are concerned here with the solution of the Maxwell equations

$$\nabla x E = ik H \qquad (2)$$

and

$$\nabla x H = -ik E, \qquad (3)$$

where H is the magnetic vector and E is the electric vector. Here $k = 2\pi/\lambda$, $i = \sqrt{-1}$ and the time dependence $\exp(-ikct)$ has been suppressed. [c is the velocity of light in units of length and time consistent with those of the space and time variables.] We seek solutions of (2) and (3) in which the x and y components of the electric field vanish on the half-plane $-\infty < x < \infty, y > 0, z = 0$, and which have as a source any one of the three fields described in (1). There are conditions at infinity which basically describe the Sommerfeld radiation condition, but we shall not give them here.

The electric field at any point in space is given by the representation

$$E(x, y, z) = E_o - \int_{sh} [ik(\boldsymbol{n}x\boldsymbol{H})\varphi$$
$$+ (\boldsymbol{n}x\boldsymbol{E})x\nabla'\phi + (\boldsymbol{n}\cdot\boldsymbol{E})\nabla'\phi] \, da',$$

where $\boldsymbol{n}$ is the unit normal vector to the plane and $\boldsymbol{E}_o$ is the vector with components (1a), (1b) or (1c) and $\phi = \exp[ik \mid \boldsymbol{r} - \boldsymbol{r}' \mid]/4\pi \mid \boldsymbol{r} - \boldsymbol{r}' \mid$. Since $\boldsymbol{n}x\boldsymbol{E}$ vanishes on the half plane, we have that

$$E_x = E_{ox} + \int_{sh} \left[ikJ_x\phi - (\boldsymbol{n}\cdot\boldsymbol{E})\frac{\partial\phi}{\partial x'} \right] da' \qquad (4)$$

and

$$E_y = E_{oy} - \int_{sh} \left[ikJ_y\phi + (\boldsymbol{n}\cdot\boldsymbol{E})\frac{\partial\phi}{\partial y'} \right] da' \qquad (5)$$

with $z' = 0$. J_x and J_y are the respective discontinuities of H_x and H_y on the half plane $-\infty < x' < \infty, y' > 0$, $z' = 0$. Eq. (4) may be transformed by an integration by parts into

$$E_x = E_{ox} + \int_{sh} \left[ikJ_x + \frac{\partial}{\partial x'}(\boldsymbol{n}\cdot\boldsymbol{E}) \right] \phi \, da' \qquad (4a)$$

[1] T. B. A. Senior, "The diffraction of a dipole field by a perfectly conducting half-plane," *Quart. Jour. Mech. Appl. Math.*, vol. 6, pp. 101–114, 1953.

[2] A. E. Heins and S. Silver, "The edge conditions and field representation theorems in the theory of electromagnetic diffraction," *Proc. Cambridge Phil. Soc.*, vol. 51, pp. 149–161, 1955.

[3] H. M. MacDonald, "A class of diffraction problems," *Proc. London Math. Soc.*, ser. 2, vol. 14, pp. 410–427, 1915.

This integration by parts is permissible since differentiation with respect to x' will not affect the order of $E \cdot n$ at the edge of the plane and $n \cdot E$ as well as ϕ approach zero at $|x'| \rightarrow \infty$, $y' \neq 0$. An integration by parts in (5) gives us

$$E_y = E_{oy} + \int_{sh} \left[ik \int_o^{y'} J_y \, d\alpha - (n \cdot E) \right] \frac{\partial \phi}{\partial y'} \, da'.$$

The additional term in the integration by parts drops out because of the facts that J_y is integrable at the origin and ϕ vanishes at infinity. In view of the fact that the y variation of ϕ appears in the form $y - y'$, this last equation may be written as

$$\int_o^y E_y \, d\beta = \int_o^y E_{oy} \, d\beta$$
$$- \int_{sh} \left[ik \int_o^{y'} J_y \, d\alpha - (n \cdot E) \right] \phi \, da' + b(x,z) \quad (5a)$$

where b is an aribtrary function of x and z.

There is a second representation which expresses the tangential components of the magnetic field in terms of the tangential components of the electric field and the normal component of the magnetic field on the plane $-\infty < x < \infty$, $y < 0$, $z = 0$. Instead of choosing ϕ with a source at $P(x, y, z) = P'(x', y', z')$, we choose a function ϕ^* whose z' derivative vanishes on the plane $z' = 0$. That is, we take

$$\phi^* = \phi(x, y, z) + \phi(x, y, -z).$$

Then for $z_o > 0$, we have for $z > 0$

$$H = H_o + H_o^r + \int_{ap} [ik(n \, xE)\phi^*$$
$$- (nxH)x\nabla'\phi^* - (n \cdot H)\nabla'\phi^*] \, da^1 \quad (6)$$

while for $z < 0$, we have

$$H = - \int_{ap} [ik(nxE)\phi^* - (nxH)x\nabla'\phi^*$$
$$- (n \cdot H)\nabla'\phi^*] \, da^1. \quad (7)$$

Here the integration is carried out over the aperture $-\infty < x < \infty$, $y < 0$, $z = 0$. The change in sign which occurs in the representation (7) is due to the change in direction of the outer normal. The vector H_o is the vector derived from E_o through the Maxwell equations, while $H_o^r = H_o(x, y, -z)$. In view of the fact that the fact that the z' derivative of ϕ^* vanishes, (6) and (7) provide us with x components;

$$H_x = H_{ox} + H_{ox}^r$$
$$+ \int_{ap} \left[ikE_y\phi^* + H_{z'} \frac{\partial \phi^*}{\partial x'} \right] dx' \, dy', z > 0 \quad (8)$$

and

$$H_x = - \int_{ap} \left[ikE_y\phi^* + H_{z'} \frac{\partial \phi^*}{\partial x'} \right] dx \, dy', z < 0 \quad (9)$$

respectively. In view of the continuity of H_x in the aperture we are left with

$$H_{ox} + \int_{ap} \left[ikE_y\phi^* + H_{z'} \frac{\partial \phi^*}{\partial x'} \right] dx' \, dy', z = 0. \quad (9)$$

An integration by parts casts (9) into

$$O = H_{oz} + \int_{ap} \left[ikE_y - \frac{\partial H_{z'}}{\partial x'} \right] \phi^* dx' \, dy', \quad (10)$$

a form which is similar to the (4a).

THE IDENTIFICATION OF THE REPRESENTATIONS

In order to find E_x, E_y and H_x at any point in space, we first examine two scalar problems. Consider a point source at the point $P_o(x_o, y_o, z_o)$ in the presence of a half-plane $-\infty < x < \infty$, $y > 0$, $z = 0$. We require the solution of the scalar wave equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2 \psi = 0 \quad (11)$$

which vanishes on the half-plane and satisfies the radiation condition. An application of Green's theorem will give us

$$\psi = F + \int_{sc} \phi \left[\frac{\partial \psi}{\partial n'} \right] da', \quad (12)$$

where $[\partial \psi / \partial n']$ is the discontinuity of $\partial \psi / \partial n'$ on the screen. This problem has been solved by MacDonald[3] almost forty years ago and we denote its solution by Ψ_1. Ψ_1 has the explicit form $I_R - I_S$ where

$$I_R = \frac{i}{2} \int_{-\mu_R}^{\infty} H_1^{(1)}(kR \cosh \mu) \, d\mu,$$

$$\mu_R = \text{arc sinh} \left\{ \frac{2\sqrt{rr_o}}{R} \cos \tfrac{1}{2}(\theta - \theta_o) \right\}$$

and $y^2 + z^2 = r^2$, $y_o^2 + z_o^2 = r_o^2$, $y_o = r_o \cos \theta_o$, $y = r \cos \theta$, $z = r \sin \theta$, $z_o = r_o \sin \theta_o$ and

$$R = \sqrt{(x - x_o)^2 + (y - y_o)^2 + (z - z_o)^2}.$$

Further

$$I_S = \frac{i}{2} \int_{-\mu_S}^{\infty} H_1^{(1)}(kS \cosh \mu) \, d\mu$$

where

$$S = \sqrt{(x - x_o)^2 + (y - y_o)^2 + (z + z_o)^2}$$

and

$$\mu_S = \text{arc sinh} \left\{ \frac{2\sqrt{rr_o}}{S} \cos \tfrac{1}{2}(\theta + \theta_o) \right\}$$

We now examine a second problem. Find the solution of (11) subject to the fact that its normal derivative vanishes on the screen and satisfies the radiation condition. Thus if $z_o > 0$, we have for $z > 0$

$$\psi = F + F^r - \int_{ap} \phi^* \frac{\partial \psi}{\partial n^1} \, da^1, \quad (13)$$

and for $z < 0$,

$$\psi = \int_{ap} \phi^* \frac{\partial \psi}{\partial n^1} \, da^1 \quad (14)$$

where $F^r = F(x, y, -z)$. Clearly, because of the properties of ϕ^*, $\partial \psi / \partial z$ vanishes on the screen. Such a function ψ may be given in terms of I_R and I_S. We call such a function Ψ_2 and we have $\Psi_2 = I_R + I_S$.

Let us now compare E_z with (11). We have that [for case 1c]

$$E_z = + \frac{\partial^2 \Psi_1}{\partial x_o \partial z_o},$$

since differentiation of (12) with respect to x_o and z_o will provide us with an equation of the same form which E_x satisfies. Similarly,

$$\int_o^y E_v \, d\beta = B(x, y, z) - \frac{\partial \Psi_1}{\partial z_o},$$

or

$$E_y = \frac{\partial B}{\partial y} - \frac{\partial^2 \Psi_1}{\partial y \partial z_o},$$

where for the present B is undetermined. Since $\Psi_1 = 0(\sqrt{r}), r \to 0, E_x = 0(\sqrt{r}), r \to 0$. Upon comparing (8) and (9) with (12) and (13) it is clear that

$$H_z = ik \frac{\partial \Psi_2}{\partial y_o},$$

and since Ψ_2 is bounded for $r \to 0$, so will H_z be bounded.

We now have enough relations to form a differential equation for $\partial B/\partial y$. The divergence condition for E gives us

$$-\frac{\partial^3 \Psi_1}{\partial x^2 \partial z_o} + \frac{\partial^2 B}{\partial y^2} - \frac{\partial^3 \Psi_1}{\partial y^2 \partial z_o} + \frac{\partial E_z}{\partial z} = 0.$$

The Maxwell equation

$$ikH = \nabla x E$$

gives us the condition (on the x component)

$$-k^2 \frac{\partial \Psi_2}{\partial y_o} = \frac{\partial E_z}{\partial y} - \frac{\partial^2 B}{\partial y \partial z} + \frac{\partial^3 \Psi_1}{\partial y \partial z \partial z_o}.$$

Upon eliminating E_z from these last two equations, we have

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\frac{\partial B}{\partial y} = \frac{\partial^4 \Psi}{\partial x^2 \partial y \partial z_o} + \frac{\partial^4 \Psi_1}{\partial y^3 \partial z_o}$$
$$+ \frac{\partial^4 \Psi_1}{\partial y \partial z_o \partial z^2} + k^2 \frac{\partial^2 \Psi_2}{\partial y_o \partial z};$$

or

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\frac{\partial B}{\partial y} = k^2\left[\frac{\partial^2 \Psi_2}{\partial y_o \partial z} - \frac{\partial^2 \Psi_1}{\partial y \partial z_o}\right]. \quad (14)$$

Eq. (14) simplifies considerably after the evaluation of the right side. We have that

$$k^2 \frac{\partial^2 \Psi_2}{\partial y_o \partial z} - k^2 \frac{\partial^2 \Psi_1}{\partial y \partial z_o} = 2 k^2 \sin \frac{\theta}{2}$$

$$\cos \frac{\theta_o}{2}\left[\frac{\partial H_1^{(1)}}{\partial R_1} \frac{(r + r_o)^2}{R_1^2 \sqrt{rr_o}} + \frac{H_1^{(1)}}{R_1 \sqrt{rr_o}} - \frac{(r + r_o)^2 H_1^{(1)}}{R_1^3 \sqrt{rr_o}}\right],$$

where $R_1 = \sqrt{(x - x_o)^2 + (r + r_o)^2}$. Since the left side of (14) is expressed in terms of y and z alone, it is clear that a measure of simplification might be attained if we express all operations in terms of r and θ. This then gives us

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right)\frac{\partial B}{\partial y}$$

$$= \frac{ik \sin \theta/2 \cos \theta_0/2}{\sqrt{rr_o}} \frac{\partial^2 H_o^{(1)}}{\partial r^2} (kR_1). \quad (15)$$

This is a Poisson equation for $\partial B/\partial y$. The solution to the homogeneous equation is inadmissable since it does not obey the Sommerfeld radiation condition. The nonhomogeneous solution may be obtained quite readily and is

$$\frac{\partial B}{\partial y} = \frac{ik \sin \theta/2 \cos \theta_o/2 H_o^{(1)}[kR_1]}{\sqrt{rr_o}}.$$

Hence E_y is determined and with the aid of the divergence condition for E we may determine E_z. Or, is we wish, $\partial B/\partial y$ may be eliminated from the differential equations at the beginning of this section and a differential equation similar to (15) will be encountered. Clearly for $r_o \neq 0, \theta \neq 0$ or $2\pi, E_y = 0(1/\sqrt{r}), r \to 0$.

A SECOND POLARIZATION

We now examine the case in which there is no component of the magnetic field parallel to the edge of the screen, that is (1a). In this case we have

$$E_x = \left(\frac{\partial^2}{\partial x_o^2} + k^2\right)\Psi_1$$

$$E_v = \frac{\partial^2 \Psi_1}{\partial x \partial y} + \frac{\partial A}{\partial y}$$

and

$$H_x = 0$$

Here A is unknown and is to be determined. As in the previous case, we employ the divergence condition on E and the Maxwell equations. From the divergence condition we obtain

$$\frac{\partial}{\partial x}\left(\frac{\partial^2}{\partial x_o^2} + k^2\right)\Psi_1 + \frac{\partial^3 \Psi_1}{\partial x \partial y^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial E_z}{\partial z} = 0.$$

From the Maxwell equations we obtain

$$\frac{\partial E_z}{\partial y} - \frac{\partial^3 \Psi_1}{\partial z \partial x \partial y} - \frac{\partial^2 A}{\partial y \partial z} = 0.$$

Upon eliminating A we obtain

$$\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)E_z = -\frac{\partial^2}{\partial x \partial z}\left(\frac{\partial^2}{\partial x^2} + k^2\right)\Psi_1$$

$$= \frac{\partial^2}{\partial x \partial z}\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\Psi_1,$$

or

$$E_z = \frac{\partial^2 \Psi_1}{\partial x \partial z} + C,$$

where C is an arbitrary harmonic function in y and z. Clearly C cannot satisfy the radiation condition and hence vanishes. From this it follows that $\partial A/\partial y$ vanishes as a consequence of the divergence condition. The remaining components of H may be determined without difficulty.

A Critique of the Variational Method in Scattering Problems[*]

D. S. JONES[†]

Summary—It is shown that the variational method of dealing with the integral equations of scattering problems is equivalent to solving the integral equation directly by Galerkin's method and using the standard formula for the amplitude of the scattered wave. The second method also satisfies the reciprocity theorem. It is therefore suggested that the reciprocity theorem be used as the basis of approximation without the introduction of variational formulas.

The error involved in using an approximate solution is discussed and it is shown that only a special set of approximations can lead to accuracy at low frequencies. Some ways in which bounds for the error may be obtained in special problems are also given.

INTRODUCTION

THE FACT that an integral equation can be formulated as a variational principle was stated as early as 1884 by Volterra[1] but it has not been used for practical calculation in scattering problems until recent years, when Schwinger discovered that the amplitude of the scattered field is closely related to the quantity whose variation has to be considered. (In connection with small-amplitude sound waves see Levine and Schwinger,[2] and in connection with atomic collisions see Schwinger.)[3] The essence of the variational method is that, if a good approximation for the field is inserted in the variational expression, an improved approximation for the scattered amplitude should result. One particular way of choosing the approximation, that given by Levine and Schwinger,[2] is to expand the field in a set of functions and solve the simultaneous linear equations resulting from the variational principle for the coefficients. It will be shown in the following that this is exactly equivalent to Galerkin's method of solving the integral equation and, furthermore, that only by expanding the field in a special set of functions can accuracy be obtained. Also this solution satisfies the reciprocity theorem.

Another way of approximation is to choose for the field an expression which is physically plausible and/or mathematically simple and insert it immediately in the variational expression. It will be shown that this is equivalent to choosing an approximation which satisfies the reciprocity theorem.

It appears, therefore, that fundamentally the use of the variational method is to ensure that the reciprocity theorem is complied with. Since the reciprocity theorem is easily written down, it is obviously simpler to formulate our analysis in terms of it than to introduce the variational principle. The advantage is mainly one of clarity for, once the equation(s) to determine the approximation have been obtained, the labor required to calculate the scattered amplitude is the same by either method.

The equivalence of the variational method and Galerkin's method[4]

The mathematical problem encountered in scattering may be stated as follows: Given a function f and operator L find a function g such that

$$f = Lg. \tag{1}$$

The function f is determined by the incident field and g can be regarded as a suitable source distribution. The equation is supposed to have a unique solution. Examples of (1) are

$$f(\mathbf{r}') = \int_S g(\mathbf{r})\Psi(\mathbf{r}, \mathbf{r}') \, d\mathbf{r}$$

for one of the boundary conditions in diffraction by an aperture, and

$$f(\mathbf{r}') = g(\mathbf{r}') + \int_T V(\mathbf{r})g(\mathbf{r})\Psi(\mathbf{r}, \mathbf{r}') \, d\mathbf{r}$$

for atomic collisions or scattering by a dielectric obstacle. In these equations

$$\Psi(\mathbf{r}, \mathbf{r}') = \frac{e^{-ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r} - \mathbf{r}'|},$$

S is the area of the aperture, T is the volume of the scatterer and V is the potential of the scattering particle. (For the dielectric obstacle V is essentially a measure of dielectric constant.)[5]

Often it is the scattered field at some distance from the obstacle and not g itself which is of primary concern. The amplitude A of the scattered field may be written

$$A = (f', g) = (g, f') \tag{2}$$

where f' is some suitable incident field and (f', g) means the inner product of f' and g. For example, the amplitude of the far field in the direction of the unit vector $\mathbf{n}$ is obtained correctly if we define (f', g) by

$$(f', g) = \int_S f'(\mathbf{r})g(\mathbf{r}) \, d\mathbf{r}$$

[1] V. Volterra, "Sopra un problema di elettrostatica," *L. Nuovo Cim.*, vol. 16, pp. 49–57; 1884.
[2] H. Levine and J. Schwinger, "On the theory of diffraction by an aperture in an infinite plane screen. I." *Phys. Rev.*, vol. 74, pp. 958–74; 1948.
[3] Quoted in W. Kohn, "Variational methods in nuclear collision problems," *Phys. Rev.* vol. 74, pp. 1763–72; 1948.

[4] Some results similar to those of this section have been obtained independently by Lippmann. See B. A. Lippmann, "Variational Formulation of a Grating Problem," N.D.A. 18–8.
[5] Eq. (1) and the following analysis are applicable to vector field problems on regarding f and g as vectors and L as a vector operator but no specific example will be quoted.

for the aperture problem or by

$$(f', g) = \int_T f'(\mathbf{r}) V(\mathbf{r}) g(\mathbf{r}) \, d\mathbf{r}$$

for the atomic collision and take $f' = e^{ikr \cdot \mathbf{n}}$. (A different choice of f' would lead to some other quantity, the near field for example, but would not invalidate the following analysis provided that it was a possible incident field.)

Let g' be such that

$$f' = Lg'. \tag{3}$$

Assume that L is a symmetric operator such that

$$(h_1, Lh_2) = (h_2, Lh_1). \tag{4}$$

Then, from (3), (4) and (1),

$$\begin{aligned}(f', g) &= (Lg', g) \\ &= (g', Lg) \\ &= (g', f)\end{aligned}$$

or

$$A = (f', g) = (f, g') = (g', Lg). \tag{5}$$

This is the reciprocity theorem which exists for most scattering problems.

By an obvious use of this theorem, we can write

$$A = \frac{(f', g)(f, g')}{(g, Lg')}. \tag{6}$$

It can be confirmed that the necessary and sufficient condition for (1) and (3) to be satisfied is that this expression for A is stationary for independent small variations of g and g' about their correct values.

Other variation principles are available (see, for example, Kohn[3]) but will not be considered here.

To demonstrate that the choice of an approximation to make (6) stationary is equivalent to Galerkin's solution of (1) we shall consider first the special case in which $f = f'$ so that $g = g'$. This corresponds, for example, to the problem of the circular aperture in a normally incident plane sound wave with the scattering observed in the direction of propagation of the incident wave. In this special case

$$A = \frac{(f, g)^2}{(g, Lg)}. \tag{7}$$

Let $G_o, G_1, \cdots$ be a set of functions each of which is independent of the others and each of which satisfy any conditions, e.g., edge conditions, imposed on g. If an infinite number of $G_o, G_1 \cdots$ is used, the set is also required to be complete. Now try an approximation G to g given by

$$G = \sum_n c_n G_n \tag{8}$$

where the sum may involve either a finite or infinite number of terms and the constants c_n are to be determined. We shall now prove the following:

Theorem 1: The necessary and sufficient conditions for the expression in (7) to be stationary for small variations of G are

$$(f, G_n) = (G_n, LG) \qquad (n = 0, 1, \dots). \tag{9}$$

G then satisfies $(f, G) = (G, LG)$ and $A = (f, G)$.

The importance of this theorem lies in the fact that

equations (9) are precisely those obtained by substituting the expansion (8) in (1) and taking the inner product of both sides with G_n, which is Galerkin's process. Then A may be calculated directly from its original definition (2) because the approximation satisfies the reciprocity theorem. Obviously this direct method of setting up the equations is to be preferred to determining them via the variation principle. It should be remarked that, if we make a one term approximation so that $G = c_o G_o$, the reciprocity theorem is satisfied if $(f, G_o) = c_o(G_o, LG_o)$ and (f, G) becomes $(f, G_o)^2/(G_o, LG_o)$ which is the same as would be calculated from (7).

We proceed now to prove the theorem. The first variation of (7) with $g = G$ when c_n becomes $c_n + \delta c_n$ is

$$\frac{\left\{ 2(f, G)\left(f, \sum_n G_n \delta c_n\right)(G, LG) - 2(f, G)^2\left(LG, \sum_n G_n \delta c_n\right)\right\}}{(G, LG)^2}. \tag{10}$$

Suppose now that equations (9) are satisfied. Then (10) vanishes if $(f, G) = (G, LG)$. Now

$$(f, G) = \left(f, \sum_n c_n G_n\right) = \left(LG, \sum_n c_n G_n\right)$$

on account of (9); hence

$$(f, G) = (G, LG) \tag{11}$$

so that the first variation is zero. Equations (9) are thus sufficient for (7) to be stationary.

Suppose now that the first variation is zero for independent δc_n. Then

$$(G, LG)(f, G_n) = (f, G)(LG, G_n) \qquad (n = 0, 1, \dots).$$

These equations are homogeneous in the constants $c_o, c_1, \dots$. Hence if there is one solution there are any number of the form $Cc_o, Cc_1, \dots$ where C is a constant. There is therefore a solution in which

$$(G, LG) = (f, G). \tag{12}$$

Such a solution satisfies (9) and also the reciprocity theorem. The necessity for the equations has therefore been shown.

On account of (11) and (12), $A = (f, G)$ and the theorem has been proved.

The corresponding theorem for the more general expression (6) will now be stated. No proof will be given since it runs parallel to that of Theorem 1. We take G' to be the approximation to g' and try

$$G' = \sum_m c_m' G_m',$$

this series containing as many terms as (8). Then

Theorem 2: The necessary and sufficient conditions for the expression (6) with $g = G$ and $g' = G'$ to be stationary for small independent variations of G and G' are

$$(f, G_m') = (LG, G_m') \qquad (m = 0, 1, \dots)$$

and

$$(f', G_m) = (LG', G_m).$$

G and G' satisfy the reciprocity relation $(f', G) = (G, LG') = (f, G')$. The value of A is either (f, G') or (f', G).

The accuracy of the approximation

To determine a good approximation it is necessary to specify precisely what is meant by good. This involves stipulating the behavior of the operators and functions in some manner. We shall choose the behavior which is appropriate to a perturbation. Attention will be concentrated on the case $f = f'$ so that $A = (f, g)$ and the reciprocity theorem takes the form $(f, g) = (Lg, g)$.

Let α be a parameter such that

(a) for some range of α

$$L = L_o + \alpha L_1 + \alpha^2 L_2 + \cdots$$
$$f = f_o + \alpha f_1 + \alpha^2 f_2 + \cdots$$
$$g = g_o + \alpha g_1 + \alpha^2 g_2 + \cdots$$
$$G = G^o + \alpha G^1 + \alpha^2 G^2 + \cdots,$$

(b) the only solution of $L_o h = 0$ is $h = 0$,

(c) $(h_1, L_p h_2) = (L_p h_1, h_2)$ for all p.

When α is small we are writing the operator L as one which is slightly perturbed from L_o. If α measures the ratio of the linear dimensions of the obstacle and the wavelength and L_o is the static operator, then we are considering the low-frequency behavior. It has been shown by Magnus[6] that assumption (a) is justified for the problem of a low-frequency plane wave falling on a circular aperture.

If we substitute the formulas of (a) in (1) and equate the coefficients of powers of α the equations to determine g are

$$\sum_{r=0}^{p} L_{p-r} g_r = f_p \qquad (p = 0, 1, \ldots). \quad (13)$$

If the inverse of the operator L_o is known g can be found, in principle, correct to any power of α. Assumption (b) ensures a unique solution.

The approximation G is to be determined so that the reciprocity theorem is satisfied, i.e., so that $(G, LG) = (f, G)$. Equating the coefficients of the like powers of α in the expansions of both sides, we obtain

$$\sum_{r=0}^{p} \sum_{s=0}^{p-r} \left(G^r, L_{p-r-s} G^s \right) = \sum_{r=0}^{p} \left(G^r, f_{p-r} \right) \quad (p = 0, 1, \ldots)$$

or

$$\sum_{r=0}^{p} \sum_{s=0}^{p-r} (G^r, L_{p-r-s} G^s - L_{p-r-s} g_s) = 0 \qquad (14)$$

on account of (13)

Now we are going to approximate (f, g) by (f, G). Both quantities can be expanded in powers of α and the coefficients of α^p are the same if

$$\sum_{s=0}^{p} (f_{p-s}, G^s - g_s) = 0. \qquad (15)$$

Consequently the first P terms of (f, g) are given correctly by the approximation if (15) holds for all $p \leq P - 1$.

Let us now suppose that $G_m = g_m (m \leq M)$. Then, in (14), we can replace G^r by g_r in all the inner products in

[6] W. Magnus, "Infinite matrices associated with diffraction by an aperture," *Quart. Appl. Math.*, vol. 11, pp. 77–86; 1953.

which $p - r \leq M$, for in these the second member is identically zero. Also G^r may be replaced by g_r in all inner products in which $r \leq M$. By this means G^r is replaced by g_r in all inner products in (14) provided that $p \leq 2M + 1$. Hence, after a use of assumption (c),

$$\sum_{r=0}^{p} \sum_{s=0}^{p-r} (L_{p-r-s} g_r, G^s - g_s) = 0 \qquad (p \leq 2M + 1)$$

or

$$\sum_{s=0}^{p} \sum_{r=0}^{p-s} (L_{p-r-s} g_r, G^s - g_s) = 0 \qquad (p \leq 2M + 1)$$

on interchanging the order of summation. This equation becomes, on taking account of (13),

$$\sum_{s=0}^{p} (f_{p-s}, G^s - g_s) = 0 \quad (p \leq 2M + 1)$$

which is the same as (15). Thus if $G^m = g_m (m \leq M)$ the first $2M + 2$ terms of (f, g) are given correctly by the approximation.

If we carry out a similar analysis with $p = 2M + 2$ we find

$$\sum_{s=0}^{2M+2} (f_{2M+2-s}, G^s - g_s)$$
$$+ (G^{M+1} - g_{M+1}, L_o G^{M+1} - L_o g_{M+1}) = 0$$

so that the $(2M + 3)$rd term of (f, g) is given correctly only if

$$(G^{M+1} - g_{M+1}, L_o G^{M+1} - L_o g_{M+1}) = 0. \qquad (16)$$

This result enables us to obtain some information which is particularly applicable to the low-frequency behavior. At low frequencies, where L_o is the static operator and α is purely imaginary, the following additional assumptions hold in most scattering problems:

(d) L_p operating on a real function produces a real function for all p;

(e) if h is real and $(h, L_o h) = 0$ then $h = 0$;

(f) f_p is purely real for all p.

With assumptions (d) and (f) it is easy to see that all g_p and G^p are real. Hence if (16) holds, assumption (e) shows that $G^{M+1} = g_{M+1}$. In particular, this applies if $M = -1$.

Consequently we have shown:

Theorem 3: The sufficient conditions for (f, G), G satisfying the reciprocity relation, to give correctly the first $2M + 2$ terms of the power series expansion of (f, g) are

$$G_m = g_m \qquad (m \leq M)$$

when assumptions (a) to (c) are satisfied. If, in addition, assumptions (d) to (f) hold, the conditions are also necessary.

The corresponding theorem for the more general case is not so powerful. It is

Theorem 4: The sufficient conditions for (f', G) or (f, G'), G and G' satisfying the reciprocity relations, to give correctly the first $2M + 2$ terms of the power series expansion of A are

$$G^m = g_m, \; G'^m = g'_m \qquad (m \leq M)$$

when assumptions (a) to (c) hold with appropriate addition to (a) for G' and g'.

The proof is similar to that for Theorem 3 and will not be given.

The low-frequency behavior

In this section we shall consider the case $f = f'$ when assumptions (a) to (f) hold. These assumptions are specially applicable to the low-frequency behavior and we shall make particular reference to the process employed by Levine and Schwinger.[2]

Assume that $G = \sum_{m=0}^{n} c_m G_m$ where the G_m are real, independent of α and defined in a similar way to those of Theorem 1. The constants c_m are determined by equations (9). One hopes to obtain an approximation which is reasonable over the whole frequency band by a suitable choice of G_m. By means of Theorem 3 we can say what this choice must be for the low-frequency end. While we cannot guarantee from this work that the choice will also be appropriate for the high-frequency end, we can say that, at any rate, it is not precluded.

Suppose that we desire that this approximation G gives the first $2n + 2N + 2$ terms of (f, g) correctly at low frequencies, N being an integer. If, in particular, $n = 0$ we want to make N as large as possible. Now, when the first $2n + 2N + 2$ terms of (f, g) are correct, it follows from Theorem 3 that $G^m = g_m (m \leq n + N)$. But G^m can be expressed in terms of $G_o, G_1, \ldots, G_n$. Hence the first $2n + 2N + 2$ terms of (f, g) can be correct only if $G_o, G_1, \ldots, G_n$ are linear combinations of $g_o, g_1, \ldots, g_{n+N}$ or, alternatively, $g_o, g_1, \ldots, g_{n+N}$ linear combinations of $G_o, G_1, \ldots, G_n$.

If, in addition, we require the first $2n + 2N + 2$ terms to be correct for all n then, on taking $n = 0$, $g_o, g_1, \ldots, g_N$ must be constant multiplies of G_o. Taking $n = 1$, we deduce that g_{N+1} must be linear combination of G_o and G_1. From $n = 2$, g_{N+2} must be a linear combination of G_o, G_1 and G_2 and the process continues.

Thus the maximum number of terms which can be obtained by a one element approximation to $G(n = 0)$ is determined by the first g_n which is not a constant multiple of g_o.

Consider now the application of these results to the problem considered by Levine and Schwinger. In that problem $f_1 = 0$, $L_1 = 0$, so that $g_1 = 0$, but g_2 is not a constant multiple of g_o. Hence $N = 1$ and the maximum number of terms of (f, g) to be obtained from the approximation $G = c_o G_o$ is 4 and can be obtained only by choosing $G_o = g_o$. From $G = c_o G_o + c_1 G_1$ we can obtain 6 terms provided that G_1 is a linear combination of g_o and g_2. Similarly, a suitable choice of G_2 will give 8 terms of (f, g) and so on. The choice of G_n made by Levine and Schwinger is in fact the one dictated by the accuracy considerations above. Of course, it is not neces-

sary to make their particular choice of G_1 to ensure low-frequency accuracy; we can choose any linear combinations of g_o and g_2 so long as both are included. Similar remarks apply to any G_n and, in general, we would choose those that are most convenient for dealing with (9).

Finally, there is one thing to be careful of in the use of Theorem 3. The statement of accuracy in this theorem depends upon L and therefore upon the accuracy with which L is approximated. Thus, when the inner products in (9) have to be evaluated approximately (as often happens), care must be taken that this evaluation does not vitiate Theorem 3. Only if the constants in (9) are obtained correct to $O(\alpha^{2M+1})$ at the low-frequency end is Theorem 3 available.

Bounds for the error

Although the analysis of the preceding section indicates how the trial function must be chosen, it does not give a precise estimate of the error nor does it state the range of α for which the approximation is good. In practice, it may well be more desirable to use an approximate field for which it is possible to say that (f, g) is estimated correctly to (say) 10 per cent for a certain range of wavelength than to insist that the field satisfies the reciprocity theorem and gives (say) 1 per cent accuracy over an unknown range. We shall now show how an estimate of the error can be made in certain circumstances.

Let $\gamma = g - G$ where G is some approximation to g. Then

$$(f, \gamma) = (Lg, \gamma) = (g, L\gamma)$$

or

$$(f, g) = (f, G) + (G, L\gamma) + (\gamma, L\gamma). \tag{17}$$

Since

$$L\gamma = f - LG \tag{18}$$

the second term may also be written as $(G, f - LG)$. It is easy to see from (17) that $2(f, g) - (g, Lg)$ is a stationary expression for (f, g).

If G satisfies the reciprocity theorem the second term in (17) is zero and the error made by the approximation is $(\gamma, L\gamma)$. This result sometimes enables us to state the sign of the error. For example, if all quantities are real and L is positive definite, we see that (f, G) is a lower bound to (f, g).

Even if G does not satisfy the reciprocity theorem the only term which cannot be calculated in (17) is the third because the second can always be evaluated by means of (18). An estimate of the goodness of the approximation therefore requires only an estimate of the third term in (17); this estimate depends essentially on being able to deduce properties of γ from a knowledge of $L\gamma$. There does not appear to be any method which we can say will be generally applicable because the operator and functions occurring are usually complex but we shall give

two methods which are available in suitable circumstances.

Firstly, suppose that it is possible to say

$$| (h_1 , h_2) |^2 \leq (h_1 , h_1{}^*)(h_2 , h_2{}^*),$$

where the star indicates a complex conjugate and suppose, further, that there is a $\mu > 0$ such that

$$(L\gamma, L^*\gamma^*) \geq \mu(\gamma, \gamma^*)$$

for all γ which can arise. Then

$$| (\gamma, L\gamma) | \leq \frac{1}{\sqrt{\mu}} (L\gamma, L^*\gamma^*).$$

Since μ and $L\gamma$ are both known this inequality gives an estimate of the error. We may also determine the accuracy of the other approximations. For $(f', g) = (f', G) + (f', \gamma)$ and

$$| (f', \gamma) |^2 \leq \frac{1}{\mu} (f', f'^*)(L\gamma, L^*\gamma^*); \tag{19}$$

all quantities are known on the right-hand side of (19).

A second method, related to that used by Kato[7] in problems of atomic collisions, depends upon adjusting the incident field until a prescribed scattered amplitude is obtained. The scattered amplitude corresponding to the original incident field can then be deduced immediately. Let ν be a complex constant such that, if

$$\nu f = Lg^\circ, \tag{20}$$
$$(f, g^\circ) = 1. \tag{21}$$

Since $g^\circ = \nu g$ we see that $(f, g) = 1/\nu$.

Consider now G chosen so that

$$(f, G) = 1. \tag{22}$$

Let $G^\circ = \nu'f - (L - 1)G = \nu'f - MG$ where ν' is a complex constant. G° is considered as an approximation to g°. Then

$$(g^\circ, G - G^\circ) = (g^\circ, LG - \nu'f) = (Lg^\circ, G) - \nu' = \nu - \nu'$$

on account of (20), (21) and (22). Hence

$$\nu = \nu' + (G^\circ, G - G^\circ) + (g^\circ - G^\circ, G - G^\circ). \tag{23}$$

We can regard the first two terms as giving an approximation to ν, the third being the error.

To estimate this error, we introduce the equation

$$(1 + \lambda)M\phi + \phi = \Lambda f \tag{24}$$

where $(1 + \lambda)(f, \phi) = 0$ and Λ is a non-zero constant to be chosen later. We assume that non-zero ϕ can be obtained only for a discrete set of real λ of which λ_n is a typical one, ϕ_n being the corresponding solution. Since

$$(\phi_n , \phi_m) = -(\lambda_m + 1)(\phi_n , M\phi_m)$$

$$= -(\lambda_m + 1)(M\phi_n , \phi_m) = \frac{1 + \lambda_m}{1 + \lambda_n}(\phi_n , \phi_m)$$

we choose Λ for each λ_n so that

$$(\phi_m , \phi_n) = \delta_{nm}$$

where $\delta_{nm} = 0(n \neq m), = 1(n = m)$. It will be assumed that the ϕ_n form a complete set.

[7] T. Kato, "Upper and lower bounds of scattering phases," *Prog. Theor. Phys.*, vol. 6, pp. 394–407; 1951.

Define the constants a_n and b_n by

$$(G - G^\circ, \phi_n) = a_n$$

and

$$(g^\circ - G^\circ, \phi_n) = b_n .$$

Then

$$b_n = (\nu f - Mg^\circ - \nu'f + MG, \phi_n)$$
$$= (M(G - g^\circ), \phi_n)$$

because of (24); thus

$$b_n = (G - g^\circ, M\phi_n)$$

$$= - \frac{1}{1 + \lambda_n} (G - g^\circ, \phi_n)$$

on account of (21) and (22). Also

$$a_n = (LG - Lg^\circ, \phi_n)$$

$$= (G - g^\circ, L\phi_n)$$

$$= \frac{\lambda_n}{1 + \lambda_n} (G - g^\circ, \phi_n)$$

$$= -\lambda_n b_n .$$

Therefore

$$(g^\circ - G^\circ, G - G^\circ) = \sum a_n b_n = - \sum \frac{1}{\lambda_n} a_n{}^2$$

and

$$(G - G^\circ, G - G^\circ) = \sum a_n{}^2.$$

If, now, we can say that a_n is real for all n, it follows that

$$- \frac{1}{\lambda_+} (G - G^\circ, G - G^\circ) \leq (g^\circ - G^\circ, G - G^\circ)$$

$$\leq - \frac{1}{\lambda_-} (G - G^\circ, G - G^\circ),$$

where λ_+ is the smallest positive λ_n and λ_- is the negative λ_n of smallest absolute value. (Note that λ_n cannot equal zero for any n.) Since G° is known in terms of ν' and G we have obtained an estimate for the error in using the first two terms of (23) as an approximation to ν. The accuracy of G° as an approximation to g° may also be obtained from

$$(g^\circ - G^\circ, g^\circ - G^\circ) = \sum b_n{}^2$$

$$\leq \max \left(\frac{1}{\lambda_+} , - \frac{1}{\lambda_-} \right)(G - G^\circ, G - G^\circ).$$

This enables us to determine the error in quantities such as (f', g°).

In practice this method may not be at all easy to apply because of the many restrictions. We have first to verify, in any particular problem, that the ϕ's satisfy the conditions imposed and then that G and ν' can be chosen so that a_n is real. (The latter condition is, at any rate, complied with when f and L are real.) Then λ_+ and λ_- (or bounds for them) must be determined. Fortunately, great accuracy is not required here because we make $(G - G^\circ, G - G^\circ)$ as small as possible by taking the best choice of ν' among those values which keep a_n real.

The Mathematician Grapples With Linear Problems Associated With the Radiation Condition

C. L. DOLPH[†]

INTRODUCTION

As the paraphrased title of von Karman's lecture (30) indicates, the complete solutions to the problems to be discussed here are not yet known. Still it seems appropriate to use the occasion of this URSI Symposium to direct attention to some recent progress as well as to some of the as yet unresolved problems connected with the mathematical foundations of several of the techniques which have been and are being employed to obtain information in scattering problems.

While a wide choice of topics would be necessary for a complete discussion, considerations of time and space make a complete survey impractical. Consequently, our discussion will be concentrated on three topics; normal mode theory, backward scattering by the Schwinger variational principle, and a "Dirichlet" principle for the wave equation. These subjects are inter-related. In the case of the first two, it is the Sommerfeld radiation condition which is the source of the difficulties while the third has been chosen as a possible way of avoiding it, or at least, as a way of transferring the difficulties elsewhere. Even in the case of these three topics, our aim will not be completeness. Rather, we shall concentrate on progress which is unlikely to have received attention as yet, either because of its recent occurrence or because of its limited availability.

THE NON SELF-ADJOINT EIGENVALUE PROBLEM OF NORMAL MODE THEORY

Very frequently in scattering problems, it becomes either convenient or necessary to treat the Sturm-Louville problem for a differential equation of the form

$$- [a(x)y']' + [b(x) - \lambda c(x)]y = 0 \qquad (1)$$

where the appropriate boundary conditions at infinity require an outgoing wave; for example, perhaps,

$$\lim_{x \to \infty} \alpha y(x) + \beta y'(x) \to 0.$$

Problems of this type already occur in Sommerfeld's book (28) on partial differential equations (cf. the Appendix devoted to a new method of determining the Green's function exterior to the sphere) and they are closely related to the Watson transformation. During the late war, much effort was devoted to the determination of the complex eigenfunctions and eigenvalues of problems of this type, as a reference to Kerr (9) will show. These problems are difficult because, in the language of the mathematician, they are not self-adjoint, and a general theory for nonself-adjoint trans-

formations does not exist. A very good idea of some of the difficulties and frustrations which result from this fact can be gained from reading Hartree et. al (7).

In fact it has been only very recently that mathematicians have realized that these problems are capable of rigorous treatment and that the type of results given by Marcuvitz (14) can be stated under precise conditions. The questions of interest include: What are the appropriate boundary conditions? Where does the spectrum lie and when will it be discrete? What are the expansion theorems and the completeness relations?

In general the methods that have been used in an attempt to obtain answers to these questions are conceptually similar although they may be phrased either in concrete or abstract language. The differential equation (1) is first transformed into the canonical form

$$-y'' + [q(x) - \lambda]y = 0 \qquad (2)$$

and a resolvent constructed for some value of λ not in the spectrum. The subsequent discussion is based on the determination of the projections associated with the resolvent via the Cauchy integral formula in either a concrete or abstract form.

The first definitive work seems to be that of Phillips (21) who showed that in the case of real $q(x)$, the classical limit point-limit circle argument of Weyl (31) could be extended to the case with boundary conditions $y(0) = 0$ and limit $x \to \infty$ $\alpha y(x) + \beta y'(x) = 0$, where the ratio α/β could be complex. In the limit point case, the second boundary condition is inessential in that only one possible interpretation exists, that of a self-adjoint problem with a real spectrum. In the limit circle case, the spectrum is discrete and lies in the upper or lower half of the complex plane depending upon whether the imaginary part of α/β is less than or greater than zero, respectively. In addition, Phillips made a start toward an expansion theory and gave an example in which the spectrum, while discrete, was not confined to any strip parallel to the real axis.

Since then, his student Sims[1] has made a thorough study of the case where $q(x)$ is complex-valued as it would be in a conducting medium. Because his results are as yet available only in his unpublished dissertation (27), we would like to indicate them in some detail. One begins as in Titchmarsh (29) by considering a finite closed interval r, b' on which $q(x)$ is complex valued and continuous with the property of having finite limits at $[r$ and $b']$. One constructs the fundamental family of solutions $\phi(x, \lambda)$ and $\theta(x, \lambda)$ to (2) with the properties that

---

† University of Michigan, on leave for 1955–1956 with University of California and the Ramo-Wooldridge Co. at Los Angeles.

$$\phi(r, \lambda) = \sin \alpha_o; \theta(r, \lambda) = \cos \alpha_o$$
$$\phi'(r, \lambda) = -\cos \alpha_o; \theta'(r, \lambda)_\lambda = \sin \alpha_o,$$

and considers the combination

$$\psi(x, \lambda) = \theta(x, \lambda) + l(\lambda, b')\theta(x, \lambda),$$

on which the boundary condition

$$\alpha\psi + \beta\psi' = 0 \text{ is imposed at } x = b'.$$

Then as in the Weyl theory [cf. Titchmarsh (11)], if $z = \alpha/\beta$,

$$l = -(\theta z + \theta')/(\varphi z + \varphi')$$

is a nonsingular linear fractional transformation which maps the lower z-plane into a circle $C_{b'}(\lambda)$ in the l-plane. Now, if b' is allowed to approach infinity,[2] a "limit-circle," "limit-point" geometry is shown to exist for $Im(\lambda) > 0$ in the sense that if $b' < b''$ then $C_b'(\lambda)$, $C_{b'}(\lambda)$ so that $C_{b'}(\lambda)$ either tends to a circle $C_\infty(\lambda)$ or a point in the l-plane. The radius of the circles $C_{b'}(\lambda)$ is given by

$$r_{b'}(\lambda) = \{-\sinh 2Im(\alpha_o) + 2\int_r^{b'} [Im(\lambda)$$
$$- Im[q(x)] \mid \varphi \mid^2 dx\}^{-1} \quad (3)$$

Now if $Im(q) \leq 0$, $Im(\alpha_o) \leq 0$, and $Im(\lambda) \geq 0$[3], if $M(\lambda)$ is a point interior to all circles, and if

$$\psi(x, \lambda) = M(\lambda)\phi(x, \lambda) + \theta(x, \lambda)$$

then ψ is a solution of (2) and the equation

$$-y'' + (q - \lambda)y = f$$

has a solution for an arbitrary $f(x)$ such that $\int_r^\infty \mid f \mid^2 dy$ exists. Furthermore,

$$\int_r^\infty \mid \psi(x, \lambda) \mid^2 [Im(x) - Im(q)] dx < \infty \quad (4)$$

Therefore, *a fortiori* it follows that

$$\int_r^\infty \mid \psi(x, \lambda) \mid^2 dx < \infty. \quad (5)$$

Three different cases are now possible at infinity, so that a schism exists between this theory and that of Weyl.

Case 1. ψ is the only square integrable solution of (2). This is a limit point case, since the radius given by (3) will go to zero.

Case 2. There exist two linearly independent solutions to (2) which are square integrable in (r, ∞), namely ψ and φ but there is just one solution ψ which is square integrable on this interval with weight function $Im(\lambda) - Im(q(x))$. Thus this is also a limit point case in that the radius defined by (3) again goes to zero.

Case 3. There exist two linearly independent solutions to (2) square-integrable on the interval $[r, \infty]$ with weight $Im(\lambda) - Im(q(x))$, hence *a fortiori*, square-integrable with a unit weighing function. This is a limit circle case since the existence of the second integral implies that the radius as defined by (3) will have a finite limit as $b' \rightarrow \infty$.

The new phenomenon is represented by Case 2. That it can actually occur is furnished by the examples

$$q(x) = -\left(x^6 - i\frac{3x^2}{2}\right) \text{ at the point } x = \infty$$

and

$$q(x) = \left(\frac{5}{10x^2} - \frac{i}{x}\right) \text{ at the singular point } x = 0$$

Furthermore as in the usual theory, these results can be extended to the double singular interval (a, b) where $a = -\infty$ and $b = \infty$ are admitted. There are then six essentially different cases. If we let the first number refer to $x = a$, the second to $x = b$, these six cases can be described as (1, 1), (1, 2), (1, 3), (2, 2), (2, 3), and (3, 3).

Of course, for the doubly singular interval (a, b) there will exist a function $m(\lambda)$ playing the same role at $x = a$ as $M(\lambda)$ does for $x = b$. For simplicity we will describe the nature of $M(\lambda)$, but, of course, corresponding statements are true about $m(\lambda)$. In Case 1, $M(\lambda)$ is uniquely determined. In Case 2, $M(\lambda)$ is a meromorphic function, and in Case 3, there exists a class of meromorphic functions $M(\lambda)$. In all cases there exists at least one analytic M function and in Case 3, an analytic M function can be found such that $M(\lambda_o) = M_o$ where M_o is an arbitrary point of an arbitrary limit circle $C_\infty(\lambda_o)$. Moreover it can be so chosen as to take all its values on the boundary of the limit circles $C_\infty(\lambda_o)$. In Cases 2 or 3, $M(\lambda)$ has a pole at λ_1, if and only if

$$1 + (\lambda_1 - \lambda)\int_r^\infty \psi(x, \lambda)\varphi(x, \lambda) dx = 0$$

for all λ. The order of this pole is one greater than that of the zero of the entire function

$$\int_r^\infty \varphi^2(x, \lambda) dx$$

at λ. Also if λ_1 and λ_2 are two distinct poles of $M(\lambda)$ then

$$\int_r^\infty \varphi(x, \lambda_1)\varphi(x, \lambda_2) dx = 0.$$

In order to describe Sims' results on the nature of the spectrum and the type of boundary conditions which are needed, it is necessary to introduce the following definitions. Let

$$\psi_o(x, \lambda) = {}^*m(\lambda)\phi(x, \lambda) + \theta(x, \lambda)$$
$$\psi_1(x, \lambda) = M(\lambda)\phi(x, \lambda) + \theta(x, \lambda)$$
$$W_x(u, v) = u(x)v'(x) - u'(x)v(x)$$
$$w(\lambda) = M(\lambda) - m(\lambda) = W_x(\psi_1, \psi_o)$$
$$R_o(\lambda) = [w(\lambda)]^{-1}\psi_1(x, \lambda)\int_a^x \psi_o(t, \lambda)f(t) dt$$
$$+ \psi_o(x, \lambda)\int_x^\infty \psi_1(t, \lambda) f(t) dt.$$

$X =$ the space of all square integrable functions on $a \leq x \leq b$.

$D^* =$ the space of all functions f in X, such that f, f' are absolutely continuous, and $(-f'' + qf)$ is in X.

[2] The same results are also true if b' is allowed to approach any finite value b for which the differential equation problem is singular.
[3] In scattering problems, $Im(q)$ and $Im(\lambda)$ satisfy this hypothesis.

$$Kf = -f'' + qf$$

D_o = the space of all functions f such that f is in D^*, and $W_\infty(\psi_1, f) = W_a(\psi_o, f) = 0$.

$L_o f = Kf$ for f in D_o

$U(\lambda) = L_o - \lambda I$

For the following definitions, let λ_o be arbitrary, $Im(\lambda_o) > 0$, and let f_o be an arbitrary element of X. Then let

$$\lambda(f, x) = [w(\lambda)]^{-1} W_\infty[f, \psi_1(x, \ )]; \quad c(f) = \int_a^\infty f f_o \, dx.$$

$$g(\lambda) = A[\psi_o(x, \lambda), \lambda_o] + C[U(\lambda_o)\psi_o(x, \lambda)]$$
$$\text{for } Im(\lambda) > 0$$

$$C(f, \lambda) = -C[U(\lambda_o)R_o(\lambda)f][g(\lambda)]^{-1}$$
$$\text{for } g(\lambda) \neq 0, \quad Im(\lambda) > 0$$

$$R_1(\lambda)f = R_o(\lambda)f + C(f, \lambda)\psi_o(x, \lambda)$$

D_1 = the set of functions f, f in D^*, such that $A(f, \lambda) + C[U(\lambda)f, \lambda] = 0$.

Then in the cases (1, 2) and (1, 3) one obtains a well-defined problem for each initial pair f_o, λ_o. Specifically, Sims proves the Theorem: Let $L_1 f = Kf$ for f in D_1, then the resolvent of L_1 is $R_1(\lambda)$; $R_1(\lambda)X = D_1$, $U(\lambda)D_1 = X$; the resolvent set of L_1 contains all of the upper half λ-plane except for zeros of the analytic function $g(\lambda)$. These zeros are in the point spectrum of L_1. Similar results, although somewhat more complicated, can be stated in the cases (2, 2), (2, 3), and (3, 3).

It should be noted that a new phenomenon has again entered this theory in that one sees from the definition of D_1 that membership in the domain of a well-defined problem ordinarily depends upon behavior of the element in question throughout the whole interval (a, b) and not just at the end points.[4] That is, the ordinary notion of the boundary value problem in the real case has disappeared. It therefore becomes important to answer the question: when will a boundary condition domain occur? That is, in general, D_1 depends on $C[U(\lambda)f, \lambda]$ which depends in turn on the behavior of f_o, f throughout the interval (a, ∞). Certainly D_1 will equal D_o if $f_o = 0$ so that membership in D_1 will then depend only on the behavior of f at $x = \infty$. In fact, for the Cases (1, 2) and (1, 3), Sims prove the following Theorem: Let D be any domain determined by λ_o and f_o. Then D will be a boundary condition domain if and only if there exists a number α such that $f_o = \alpha\psi_o(x, \lambda_o)$.

Again similar though more complicated results can be stated in the cases (2, 2), (2, 3) and (3, 3).

In addition, in these last three cases, Sims was able to prove that the spectrum of L_o was a pure point spectrum consisting of discrete points in the lower half λ-plane.

It is easy in some cases to see heuristically why such a "definiteness" property must hold for the discrete

eigenvalues. Thus if λ, $y(x, b)$ is a solution of the eigenvalue problem

$$-y'' + q(x)y = \lambda y$$
$$y(a) = 0$$
$$y'(b) - iy(b) = 0 \qquad (6)$$

[a radiation condition corresponding to traveling waves of the form $\exp i(x - at)$], then it follows that

$$-\int_a^b yy'' \, dx + \int_a^b q(x) |y|^2 \, dx = \lambda \int_a^b |y|^2 \, dx$$

and

$$\int_a^b yy'' \, dx = yy' - \int_a^b |y'|^2 \, dx$$
$$= i |y(b)|^2 - \int_a^b |y'|^2 \, dx.$$

Thus, taking the imaginary part of both sides of this equation, we have

$$-|y(b)|^2 + \int_a^b Im[q(x)] |y|^2 \, dx$$
$$= Im\lambda \int_a^b |y|^2 \, dx. \qquad (7)$$

Consequently, if $Imq(x) \leq 0$ for all x, $a \leq x < \infty$ and if the radiation condition (6) is valid in the limit as $b \to \infty$, then (7) implies that $Im(\lambda) \leq 0$.

While the work of Sims is definitive in regard to the question of boundary conditions and type of spectra that can be expected under a given set of such conditions, many questions concerning expansion theorems, convergence criteria, and completeness relations are still open. Most results so far obtained in this direction have been confined to theorems relating to a finite interval (a, b). The basic difficulty consists in finding suitable asymptotic estimates for the eigenvalues, no easy task in general, as the example of Phillips referred to earlier clearly indicates. Thus, for example, if we define our domain of definition by

D^{**} = the set of functions f, f in D^* such that

$$W_a(f, \phi) = W_b(f, \psi) = 0$$

where

$$\varphi(a, \lambda) = \sin \alpha_o; \quad \psi(b, \lambda) = \sin \beta_o$$
$$\varphi'(a, \lambda) = -\cos \alpha_o; \quad \psi^1(b, \lambda) = -\cos \beta_o$$

and assume that $q(x)$ remains bounded at $x = a$ and $x = b$, then Sims was able to show that there exist at most a finite number of zeros of the Wronskian of the resolvent of multiplicity greater than one; that if $q(x)$ and α are given in the boundary condition, then this Wronskian can have zeros of order greater than one if and only if the other parameter β of the boundary condition is chosen from a certain countable set whose limit points are a discrete set contained among the points 0, $\pm n$ in the complex β-plane. Moreover, for any $\xi > 0$, one can find an $N(\xi)$ such that each open rectangular region $-\xi < \nu < \xi$, $(b - a)^{-2}(n + \frac{1}{2})^2\pi^2 < \mu < (b - a)^{-2}(n + \frac{3}{2})^2\pi^2$, $\lambda = \mu + i\nu$, contains one and only one λ_n, and each λ_n is so contained for $n > N(\xi)$.

---

[4] The physical significance and occurrence of these global domains should be investigated further.

From this, one can prove the following theorem about weak convergence. Let f be in D^{**} and let g be any element of X. Then

$$\int_a^b fg\,dx = \sum C_n(f)C_{\hat{n}}(g)$$

where

$$C_n(f) = \int_a^b \varphi(x, \lambda_n)f(x)\,dx,$$

and where λ_n is a zero of the Wronskian of the resolvent. One can further prove a completeness relation which states that if f is in D^{**} (so that f is in X) then

$$\int_a^b |f|^2\,dx = \sum_1^\infty C_n(f)C_n(f),$$

where C_n is defined as before but where it should be noticed that

$$C_n(f) \neq C_n(f).$$

Finally, if f is in the domain of definition of the resolvent then the series

$$\sum C_n(f)\varphi_n(x)$$

will converge uniformly but not absolutely to $f(x)$ as long as the boundary conditions are not chosen from the exceptional set leading to multiple roots of the Wronskian of the resolvent.

Similar results obtained by more abstract methods are due to Schwartz (24). Since these have already been reviewed in detail (in *Math. Rev.*), we will not comment further on them here. The methods of Schwartz have been modified and supplemented by Kramer (10) to yield similar results under more general conditions. Finally, for the finite interval, attention should be called to the report by Friedman and Mishoe (5) who obtained interesting results when $f(x)$ was a function of bounded variation.

The real case of interest, however, in scattering theory is that of the semi-infinite interval. Here the only treatment of expansion theorems seems to be that of Naimark (18), which, while very complete, is unfortunately limited to cases where the discrete spectra consist of a finite number of points. Naimark assumes that this will be the case by requiring either that the following integral will exist for some $\epsilon > 0$;

$$\int_0^\infty \exp(\epsilon x)\,|\,q(x)\,|\,dx < \infty,$$

or that the integral

$$\int_0^\infty x^2\,|\,q(x)\,|\,dx$$

exists. Under these circumstances the continuous spectrum is furthermore confined to the positive real λ axis. If one sets $\lambda = s^2$, lets θ be a complex parameter, and denotes in the usual fashion the asymptotic solutions as follows:

$$y_1(x, s) \sim \exp(isx) \quad \text{for } Im(s) \geq 0 \text{ as } x \to \infty$$
$$y^*(x, s) \sim \exp(-isx) \quad \text{for } Im(s) \leq 0 \text{ as } x \to \infty$$

and if

$$A(s) = y'(0, s) - \theta y_1(0, s) \quad \text{for} \quad Im(s) \geq 0$$
$$A^*(s) = y^{*'}(0, s) - y^*{}_1'(0, s) \quad \text{for} \quad Im(s) \leq 0$$
$$y(x, s) = A^*(s)y_1 - A(s)y^*{}_1 \quad \text{for} \quad Im\,s \leq 0,$$

$y_k(x) = y(x, s_k)$ for s_k or responding to an eigenvalue, then under these circumstances the resolvent is a bounded integral operator whose kernel $K(x, \zeta, \lambda)$ can be represented by

$$K(x, \zeta, \lambda) = \sum_1^r \frac{y_k(x)y_k(\zeta)}{(\lambda_k - \lambda)\int_0^\infty y^2{}_k\,dx}$$

$$- \frac{1}{2\pi}\int_0^\infty \frac{y(x, s)y(\zeta, s)\,ds}{(s^2 - \lambda)A(s)A^*(s)}$$

for any λ not in the spectrum, where the integral converges uniformly and absolutely. Similarly, one obtains an expansion theorem of the form

$$y(x) = \sum_1^r \frac{\alpha_k y_k(x)}{\int_0^\infty y_k^2\,dx} = \frac{1}{2\pi}\int_0^\infty \frac{\alpha(s)y(x,s)}{A(s)A^*(s)}\,ds \quad (8)$$

where

$$\alpha_k = \int_0^\infty g(x)y_k(x)\,dx \quad (9)$$

$$\alpha(s) = \int_0^\infty g(x)y^*(x, s)\,dx \quad (10)$$

when g is integrable on the interval $(0, \infty)$, g' exists and is absolutely continuous, $(-g'' + qg)$ is square-integrable on $(0, \infty)$, and $g'(0) = \theta g(0)$. Also for functions g satisfying the conditions just stated and for an arbitrary square integrable function h on the interval (O, ∞), he finds

$$\int_0^\infty g(x)h(x)\,dx = \sum_1^r \frac{\alpha_k \beta_k}{\int_0^\infty y^2{}_k\,dx}$$

$$- \frac{1}{2\pi}\int_0^\infty \frac{\alpha(s)\beta(s)}{A(s)A^*(s)}\,ds \quad (11)$$

where α_k, β_k, $\alpha(s)\beta(s)$ are defined similar to (9) and (10). In (8) the integrals converge absolutely and uniformly in x while in (11) they converge absolutely. Naimark also discusses similar theorems when a degeneracy can occur in the finite discrete spectra.

In view of recent progress, it is to be hoped that it will not be too long a time before the remaining gaps in our knowledge of problems of this type will be filled.

THE VARIATIONAL PRINCIPLES OF MacFARLANE AND OF SCHWINGER

In 1944 MacFarlane (13) observed that the method of Rayleigh-Ritz could be employed to determine the complex eigenvalues of the normal mode problem. Subsequently, the same method was employed by Pekeris and Ament (20) for the same purpose. The fact that a variational principle exists for this nonself-adjoint eigenvalue problem is remarkable and may be explained by observing that while the problem is nonself-adjoint in a Hilbert space with a Hermitian inner product, it is

formally self-adjoint in a vector space with a symmetric inner product. Thus, as has already been observed by Sommerfeld (28) and as we have exhibited it here explicitly in the case of Sims expansion theorem, the complex-valued eigenfunctions are orthogonal in the sense

$$\int_a^\infty y_k(x)y_l(x)\, dx = \delta_{kl}.$$

This leads to the conclusion that the Rayleigh quotient

$$R(y) = \frac{\int_a^\infty y[-y'' + q(x)y]\, dx}{\int_a^\infty y^2\, dx} \tag{12}$$

is stationary at an eigenvalue of the normal mode problem, where $y(a) = 0$.

Somewhat more generally, one can consider making

$$R(y, z) = \Big\{ \int_a^\infty (y'z' + qyz)\, dx$$
$$- \lim_{b \to \infty} iz(b)y(b) \Big\} \Big/ \int_o^\infty yz\, dz \tag{13}$$

stationary as a function of y and z. From the identities

$$\int_a^b (y'z' + qyz)\, dx - iy(b)z(b) = z(y' - iy)\big|_b\,;$$

$$\int z[-y'' + q(x)y]\, dx = y(z' - iz)\big|_b$$
$$+ \int y[-z'' + qz]\, dx,$$

one readily sees that in addition to obtaining

$$-y'' + qy = \lambda y$$

for functions satisfying $y(a) = 0$, one also obtains the adjoint problem as a necessary condition for (13) to be stationary:

$$-z'' + qz = \lambda f;\ \lim_{x \to \infty} (f' + if) = 0.$$

In both cases the identities given above make it clear that the radiation conditions enter as natural boundary conditions.

During these same war years Schwinger had developed his now famous variational principle and successfully applied it first to the interior problems associated with obstacles in wave guides [cf. Saxon (23)] and then to a whole series of exterior scattering problems, in collaboration with Levine (25). In the simple scalar case, that of backward scattering from a plane wave incident on a closed obstacle, it is well known that the stationary value of an expression of the form

$$S(\varphi) = \frac{(\int f\varphi\, ds)^2}{\int\int k(s, t)\varphi(s)\varphi(t)dsdt} \tag{14}$$

is proportional to the value of the complex amplitude function of the backward scattered wave, and where the kernel is $(\exp ikr)/r$ for the boundary condition of total zero scalar potential on the obstacle.

In the most general Scwhinger case, we desire to make

$$\frac{(a, \varphi')(a', \varphi)}{(k\varphi, \varphi')} \tag{15}$$

stationary where in an appropriate inner product we have that

$$(k\varphi, \varphi') = (\varphi, k'\varphi').$$

The necessary conditions for the stationary value lead as in the second formulation of MacFarlane's principle to the adjoint equations [cf. (15, 16)]

$$a = k$$

and

$$a' = k'.$$

For the sake of simplicity we will confine our attention to the symmetric formulations (12) and (14). Even with this restriction both principles have the disadvantage that little is known about the nature of their stationary points for exterior problems. Thus, while it is classical that the Rayleigh quotient has a minimum for its stationary value for many self-adjoint problems, and while necessary and sufficient conditions are known for the Schwinger principle [usually in reciprocal form to that stated in (14)] to take on a minimum for its stationary value [cf. Saxon (23) and Kato (8)], the complex-valuedness of the functions involved in (12) and in (14) in connection with a symmetric rather than a Hermitian inner product leave the nature of their stationary points in doubt.

Until these open questions have been answered, these principles are of value only as a relatively easy way of obtaining engineering answers, the validity or reasonableness of which must then in turn be verified either by comparison with an exact solution or by experiment.

In the case of MacFarlane's principle there is the hope that the "definiteness" property of the eigenvalues discussed in the first section can be used to provide a partial answer to these questions. Unfortunately, a corresponding property does not appear to be available for the Schwinger Principle in general.

In the case of MacFarlane's principle, a first step in understanding the nature of the stationary point was taken by Dolph, Marx, McLaughlin (3) in that they were able to give a characterization of the imaginary part of the eigenvalues, for a finite dimensional analog of (12) when the eigenvalues were all confined to one half of the complex λ-plane. That is, if we allow ourselves to extrapolate from our present knowledge of expansion theorems and completeness relations, we would expect that at least in Sims' cases (2, 2), (2, 3) and (3, 3) there would be theorems similar to those he derived for the finite interval. In that case, if y were an approximating function, it could be expanded in the series of complete (symmetric) orthogonal eigenfunctions $y_k(x)$

$$y = \sum_1^\infty z_i y_i(x)$$

so that (12) would reduce to

$$R(y) = R(z_1, z_2, \cdots) = \Big(\sum_1^\infty \lambda_i z_i^2 \Big) \Big/ \Big(\sum_1^\infty z_i^2 \Big). \tag{16}$$

The property of rendering (16) stationary is, moreover, equivalent to the real problem of making the imaginary

part of the numerator of the Rayleigh quotient stationary subject to the requirement that the denominator be normalized. That is, it is equivalent to making

$$I(z_1, z_2, \cdots) = \sum_{k=1}^{\infty} \nu_k(x_k^2 - y_k^2) + 2\mu_k x_k y_k$$

stationary subject to

$$\sum_{1}^{\infty} x_k^2 - y_k^2 = 1, \qquad (17)$$

and

$$\sum_{1}^{\infty} x_k y_k = 0 \qquad (18)$$

where $z_k = x_k + iy_k$ and $\lambda_k = \mu_k + iv_k$.

In any finite approximation, the requirement that $\nu_k < 0$ or that $\nu_k > 0$ enables one to characterize the $\nu_k = \text{Im}(\lambda_k)$ successively as "saddle points" in appropriately chosen orthogonal subspaces. While this characterization is not stated by Dolph *et al.* (3) in a suitably invariant form for immediate application to the infinite dimensional situation, put in a suitable form it would lead to an error estimate from one side for the imaginary parts of the complex eigenvalues of the normal mode theory. It would differ from the corresponding estimates obtained by the intrinsic maximum-minimum characterization of real eigenvalues for the Rayleigh quotient in that even the first ν_k would be a maximum-minimum and the others would be similarly obtained successively rather than by an intrinsic process.

Even this much seems beyond reach in the case of Schwinger principle. In order to state our reasons for believing this to be the case, we will proceed in a manner analogous to that recently used by Chambers (1) in his discussion of the relation between the Rayleigh-Ritz principle and that of Schwinger in the real positive-definite case. Let us imagine that we have a set of functions (ψ_k) complete for the Schwinger case. We think of them as orthogonalized with respect to the kernel $k(s, t)$ in the sense that if $\chi_k(s) = \int k(s, t)\psi_k(t)\, dt$ then

$$\int \chi_k(s)\psi_l(s)\, ds = \delta_{kl}.$$

Thus, if φ is any approximating function, we may think of it as expanded as $\varphi = \sum_{1}^{\infty} z_k \psi_k(s)$ with coefficients z_k given by $z_k = \int \varphi(s)\chi_k(s)\, ds$. Similarly we may think of the known function f as expanded as

$$f = \sum_{1}^{\infty} f_i \chi_i(s),$$

with coefficients given by

$$f_k = \int f(s)\psi_k(s)\, ds.$$

If these expansions are inserted into (14), it becomes

$$S(\varphi) = S(z_1, z_2, \cdots) = \left(\sum_{1}^{\infty} f_i z_i\right)^2 / \sum_{1}^{\infty} z_i^2. \quad (19)$$

Now if we introduce any complex rotation in which

$$z_1' = \sum \frac{f_i}{\sqrt{\sum f_i^2}} z_i$$

and where we choose $z_2', z_3', \cdots$ to preserve

$$\sum_{1}^{\infty} (z_k')^2 = \sum_{1}^{\infty} f_k^2$$

as may always be done, then (19) can be written in the form

$$S(z_1, z_2, \cdots) = [\omega(z_1')^2]/[\sum_{1}^{\infty}(z_k')^2]; \omega = \sum_{1}^{\infty} \epsilon f_i^2 \quad (20)$$

which is easily recognized as a highly degenerate form of (16) in which only one eigenvalue, ω, occurs. The problem of making (20) stationary is equivalent to the real problem of making either the real or imaginary part of the numerator of (20) stationary subject to a normalization of the denominator. If again we use the imaginary part letting $\omega = \sigma + i\tau$, $z_k = x_k + iy_k$, we must make

$$\tau[(x_1')^2 - (y_1')^2] + 2\sigma x_1' y_1' \quad (21)$$

stationary subject to (17) and (18). Eliminating x_1' and y_1' by these last two relations leads to the free problem of making

$$\tau\left\{1 - \sum_{2}^{\infty}[(x_k')^2 - (y_k')^2]\right\} - 2\sigma \sum_{2}^{\infty} x_k' y_k'$$

stationary. The solution to this problem is $x_k' = y_k' = 0$, if $|\omega|^2 = \sigma^2 + \tau^2 \neq 0$, $k = \pm 1, 2, \cdots$. Thus from (17) and (18) we find $x_1' = +1$, $y_1' = 0$ so that the stationary value of (21) is τ, leading to the stationary value ω of (20). Although as in (16), this stationary point is again a "saddle", we do not know how the "saddle" is oriented since we have no *a priori* knowledge about τ or σ, the quantities we are trying to determine; in contrast to the situation with regard to MacFarlane's principle where we know that all the ν_k are, for example, positive. A similar situation obtains if we use the real part of (19).

Thus, while, in the case of MacFarlane's principle, it is immediately possible to reject all approximating functions which lead to approximate eigenvalues whose imaginary parts are of the wrong sign as well as to hope to obtain a one-sided estimation for these imaginary parts, in the case of the Schwinger principle, nothing similar appears possible and we may be above or below the true real and imaginary parts of the stationary value in a completely arbitrary way and these will probably differ from approximation to approximation.

The situation is further complicated for both principles in that expansion of trial functions in terms of a complete set of known functions is probably best interpreted as a bi-orthogonal expansion in a Hilbert space with Hermitian inner product as can be seen from (13) or (15). While we can, at least in some instances, expect Parseval relations, as the work described in the first section indicates, there appears to be no analog of Bessel's inequality to insure that the remainder in any finite approximation goes to zero in a uniform fashion. It is not even known under what circumstances, if any, one can state a generalization of Bessel's inequality similar to that to be found on page 206 of Riesz-Nagy (22).

So far our discussion has been limited to the case of backward scattering. The situation is even more obscure for arbitrary angles of incidence and reflection. In fact, there appears to be no analogs to the "definiteness" property associated with MacFarlane's principle in any of the Schwinger cases except that of forward scattering where the familiar relation between total cross section and the imaginary part of the transmitted amplitude is encountered [cf. Morse and Feshbach (16)]. In the scalar case with $(\exp ikr)/r$ as the kernel, this "definiteness" property can also be derived more directly by use of Lemma 3 of Müller (17).

Nevanlinna's Dirichlet Principle

In our discussion up to this point, we have emphasized some of the difficulties which result from the complex radiation condition as well as some recent progress in coping with them. In this section we should like to discuss still a different approach which, in one form, avoids the complex radiation condition altogether, and which, in another form, offers considerable promise of furthering our understanding in its presence.

In a series of papers beginning in 1952, Nevanlinna (19) · began the development of a general theory of indefinite self-adjoint forms in Hilbert space with a view of deriving a "Dirichlet Principle" applicable to the wave equation with Cauchy initial data as well as to interior boundary value problems for elliptic partial differential equations such as the scalar wave equation. Unfortunately, to date the only detailed application of his methods has been made by his student, Louhivaara (12), who treated the second class of problems mentioned above. For these problems the theory of Nevanlinna is not really necessary since they lead to problems involving semi-bounded operators in Hilbert space and hence can be treated by, for example, the methods of Friedrichs (6). Nevertheless we will describe briefly both the general ideas underlying Nevanlinna's work as well as the application by Louhivaara in the hope of attracting other research workers to its interesting possibilities.

In the case of an interior boundary value problem for Laplace's equation, say

$$\nabla^2 u = 0 \text{ in a domain} \tag{22}$$

and

$$u = f \tag{23}$$

on the boundary, Dirichlet's principle says in effect that the solution may be obtained by minimizing the Dirichlet distance between f, considered as extended to the interior of the domain, and the linear space V of all functions which vanish on the boundary of the given domain.

That is, the solution is given by $u = f - v$, where v is the function in V rendering the integral

$$D(v) = \iint (f_x - v_x)^2 + (f_y - v_y)^2 \, dx \, dy \tag{24}$$

a minimum. In addition,

$$D(u, v) = \iint (u_x v_x + u_y v_y) \, dx \, dy \tag{25}$$

may be considered as a scalar product of the functions u, v, and from the identity

$$\iint (u_x v_x + u_y v_y) \, dx \, dy = \int v \frac{\partial u}{\partial \eta} \, dr - \iint \nabla^2 u \, dx \, dy \tag{26}$$

it follows that the solution of the boundary value problem (22) and (23) is orthogonal in this sense to the linear space V.

Some of these geometrical properties are preserved in more general problems but not all. Thus for the corresponding problem of the scalar wave equation in an interior domain where, for real k,

$$(\nabla^2 + k^2)u = 0 \tag{27}$$

and where

$$u = f \tag{28}$$

on the boundary, the identity corresponding to (26) is

$$\iint (v_x u_x + v_y u_y - k^2 vu) \, dx \, dy$$
$$= \int v \frac{\partial u}{\partial \eta} \, ds - \iint v(\nabla^2 + k^2)u \, dx \, dy \tag{29}$$

which shows that the orthogonality property is preserved but the quadratic form corresponding to (24) is

$$\iint (v_x^2 + v_y^2 - k^2 v^2) \, dx \, dy \tag{30}$$

which is no longer positive-definite so that the notion of "distance" loses its immediate meaning.

Similarly for the wave equation-$u_{tt} - u_{xx} = 0$—we use the one dimensional form for simplicity—the relations corresponding to (26), 25 are

$$\iint (v_t u_t - v_x u_x) dx \, dt$$
$$= \int vu_x \, dt + \oint v \, u_t \, dx - \iint v(u_{tt} - u_{xx}), \, dx \, dt \tag{31}$$
$$\iint (v_t^2 - u_x^2) dx \, dt. \tag{32}$$

Thus, if in this case we can determine an appropriate domain on the boundary of which it is reasonable to require that $v = 0$, once again we will have an orthogonality property. Although he has not yet published any results in this direction, Nevanlinna indicated in a lecture given at the University of Michigan in the summer of 1953 that he believed he knew how correctly to determine this domain in order to permit the application of the general theory we now sketch.

In each of these last two cases, it is easy to obtain a majorant for the indefinite quadratic forms involved. For example we note that for (3.0)

$$\left| \iint (|v_x|^2 + |v_y|^2 - k^2|v|^2) dx \, dy \right|$$
$$\leq \iint (v_x^2 + v_y^2 + k^2 v^2) dx \, dy \tag{32}$$

while for (31)

$$\left| \iint (v_t^2 - v_x^2) dx \, dt \right| \leq \iint (v_t^2 + v_x^2) dx \, dt \tag{33}$$

These in turn induce majorants for the associated bilinear forms (29) and (31). More generally, then, Nevan-

linna presupposes that he has a self-adjoint bilinear form $Q(u, v)$, $Q(v, u) = Q(u, v)$ which can be majorized by a positive-definite form $E(u, v)$, self-adjoint again in that $E(v, u) = E(u, v)$, and with the property that $|Q(u, v)| \leq E(u)E(v)$.

Let us now think of $E(u)$ as the unit form in a possible Hilbert space. In our examples, E would be given by (24) for the Dirichlet problem and by the right-hand sides of (32) and (33) respectively. If this can be done, then for fixed v, $Q(u, v)$ is a bounded linear functional and as such it can be written as an inner product of u and some element g of the Hilbert space:

$$Q(u, v) = E(u, g) \qquad (34)$$

[cf. for example, Riesz-Nagy (22)]. If we now allow v to vary, then g will vary linearly with v and we may write $g = g(v)$. It turns out, as is well known, that $g(v)$ is a linear bounded self-adjoint transformation whose spectrum is confined to the real λ interval $-1 \leq \lambda \leq 1$. (The self-adjointness of g follows from that of the forms Q and E). If the spectral decomposition of g can be found, then by (34) this decomposition will induce a spectral decomposition of the form $Q(u, v)$. In particular, note that if λ, v are eigenvalue and eigenfunction of g, $g(v) = \lambda v$, then (34) reduces to

$$Q(u, v) = \lambda H(u, v) \qquad (35)$$

for all u, which we recognized as a necessary condition that λ, v are eigenvalue and eigenfunction respectively of the form Q—that is, λ is a stationary value of $Q(u)/E(u)$. Furthermore the spectral decomposition of the induced operator g enables one to project an arbitrary vector in the Hilbert space into three orthogonal spaces, the space spanned by the projections associated with the negative spectrum of g, the space spanned by the projections associated with the positive spectrum of g, and the residual space associated with $\lambda = 0$ on which, from (35), $Q(u, v) = 0$. If this residual space is empty, then each element in the Hilbert space has a unique Fourier decomposition in terms of the above decomposition. Of course in this brief sketch, we have left much unsaid, concerning such things as: When can this be done? When is the result independent of the majorant? etc. The reader is referred to Courant-Hilbert, (2) for details in the positive definite case and to the above quoted works by Nevanlinna.

In order to render these abstract ideas somewhat more concrete, consider the problem formulated above for the scalar wave equation. For this problem, (34) becomes

$$\iint (u_x v_x + u_y v_y - k^2 uv)dx\,dy$$
$$= \iint [u_x g(v)_x + u_y g(v)_y + k^2 ug(v)]dx\,dy.$$

This we may transform by integration by parts to the form

$$\oint u \frac{\partial v}{\partial \eta} ds - \iint u(\nabla^2 + k^2)vdx\,dy$$

$$= \oint u \frac{\partial g}{\partial \eta} ds - \iint u(-\nabla^2 + k^2)g(v)dx\,dy.$$

Thus, if we use the standard technique of this theory, [cf. Courant-Hilbert (2)] and introduce a space of functions which vanish in a strip near the boundary of our given domain, as well as the space D_o consisting of those functions which are limits of sequences of functions from our previous space in the sense that

$$H(\varphi_n - \varphi) = \iint (\varphi_n - \varphi)^2 dx\,dy \to 0; D(\varphi_n - \varphi)$$
$$= \iint \{ (\varphi_n - \varphi)_x{}^2 + (\varphi_n - \varphi)_y{}^2 \} dx\,dy \to 0,$$

then our operator $g(v)$ will be defined by

$$\iint u(-\nabla^2 + k^2)g(v)dx\,dy = \iint u(\nabla^2 + k^2)vdx\,dy,$$

for all functions u in D_o. To determine the spectral decomposition of $g(v)$ we therefore consider the eigenvalue problem corresponding to (35):

$$Q(u, v) = \iint (u_x v_x + u_y v_y - k^2 uv)dx\,dy$$
$$= \lambda \iint (u_x v_x + u_y v_y + k^2 uv)dx\,dy$$
$$= \lambda[Q(u, v) + 2k^2 H(u, v)]$$

for all u in D_o. Now this can be transformed to the form

$$Q(u, v) = \frac{2k^2\lambda H(u, v)}{1 - \lambda},$$

or, if we add $k^2 H(u, v)$ to both sides of this last relation, to the form

$$D(u, v) = Q(u, v) + k^2 H(u, v) = \mu H(u, v), \qquad (36)$$

where

$$\mu = k^2 \frac{(1 + \lambda)}{1 - \lambda}$$

or,

$$\lambda = \frac{\mu - k^2}{\mu + k^2}; \qquad (37)$$

this verifies the fact that λ is confined to the interval $(-1 \leq \lambda \leq 1)$. This last problem however is in turn equivalent to the eigenvalue problem for the differential equation

$$v_{xx} + v_{yy} + \mu v = 0,$$

whose corresponding Rayleigh quotient is

$$R(v) = \frac{D(v)}{H(v)} = \frac{\iint (v_x{}^2 + v_y{}^2)dx\,dy}{\iint v^2 dx\,dy},$$

where $D(v)$, $H(v)$ are positive definite. Thus, the results of Chapter VII of Courant-Hilbert (2) guarantee the existence of a denumerable spectrum, $\mu_1 \leq \mu_2 \leq \cdots$ with associated eigenfunctions v_1, v_2, $\cdots$. If we denote by μ_n the value corresponding to λ_n as given by (37), and if for simplicity we assume that $k^2 \neq \mu_n$, $n = 1, 2, \cdots$ then if ψ is any function in D_o, we may think of it as expanded in these eigenfunctions:

$$\psi = \sum C_n v_n , C_n = \iint \psi v_n \, dx\,dy = H(\psi, v)$$

so that (36) implies

$$Q(u, u) = \sum_{\mu n > k}^{\infty} (\mu_n - k^2) \, | \, C_n \, |^2$$

$$+ \sum_{\mu n < k}^{\infty} \widetilde{}(\mu_n - k^2) \, | \, C_n \, |^2. \qquad (38)$$

Thus, to solve the first boundary value problem for the scalar wave equation, we define $v = u - f$ (where we again think of the boundary function f as extended to the whole domain), so that v is in D_o, and make $Q(f - v)$ stationary. By (38), we can first project $(f - v)$ into the space of eigenfunctions associated with $\lambda_n > 0$. In this subspace $Q(f - v)$ is bounded from below and the usual processes of the direct methods of the calculus of variations will guarantee the existence of a minimizing function, say $(v+)$. We can then project $(f - v)$ into the space of eigenfunctions associated with $\lambda_n < 0$. In our case this is a finite dimensional subspace and there is no difficulty in seeing that $-Q(f - v)$ will again have a minimum here, say for $(v-)$. The solution of the boundary value problem will then be given by $u = (v+) + (v-) + f$, and the stationary value of $Q(f - v)$ is clearly seen to be given by the difference of the minima associated with the positive and negative eigenvalues λ_n. Finally if $k^2 = \mu_n$ for some n, then $\lambda_n = 0$ is an eigenvalue of $Q(u, v) = 0$ and we must consider the space spanned by the eigenfunctions associated with it. In this case, the solution to the boundary value problem is unique up to an arbitrary additive function from this space, if the problem has a solution at all. A necessary and sufficient condition for it to have a solution is that $Q(f, \varphi) = 0$ for all functions φ in the space spanned by the eigen functions associated with $\lambda = 0$.

While the above theory is adequate to handle interior problems it must be extended for exterior problems. If a similar program can be carried through for the time dependent wave equation, as Nevanlinna has indicated it can, the results will be most interesting. Certainly one can anticipate that the spaces associated with the positive and negative values of λ will both be infinite dimensional and for an exterior problem we will be faced with a continuous spectrum in all probability. On the other hand, this approach avoids the complex radiation condition altogether and substitutes for it the inconvenience of a knowledge of initial conditions. Unfortunately, I have not been able to see even formally exactly how this theory would work in this case because of the difficulties associated with the domains of dependence of hyperbolic equations. Moreover, while the Sommerfeld radiation condition does cause many mathematical difficulties, initial conditions are also unknown in most problems of physical interest. In view of recent progress, some of which we have described in the first section, in understanding nonself-adjoint problems, it might be more useful to attempt a generalization of Nevanlinna's procedure for these problems. I would therefore like to conclude this report by a few preliminary notions as to how this might be accomplished.

For the sake of simplicity we will limit our discussion to consideration of the boundary value problem for the scattered potential. However, should the idea we are about to state prove fruitful, it could equally well be applied for the determination of the scattered amplitude as indicated by Levine in part II of (11).

The basic representation theorem employed by both Nevanlinna and Friedrichs (34) may be applied whenever the bilinear form $Q(n \cdot v)$ appropriate to a given problem can be majorized by a self-adjoint quadratic form $E(\mu)$. However, if a quadratic form $Q(u, v)$ is non self-adjoint, that is, if $Q_1(v, u) \neq Q_1(u, v)$, the transformation $g(v)$ will not be self-adjoint and we cannot, therefore, be sure that there will be a spectral decomposition associated with it. If there is, then the spectrum will be confined to the unit circle in the λ plane. It is my conjecture that in a great many cases of physical interest the non-self adjoint linear transformation defined by (34) will fall into Dunford's (4) classification of "spectral" transformations and that it will therefore be possible to use its spectral resolution to discuss the more general type of stationary point associated with such a non-self adjoint Q. Whether the general situation will require the consideration of the four spaces associated with the quadrants of the spectral circle or whether some definiteness property such as occurs for MacFarlane's principle will cut this number down to two, is not clear. The answer to this and similar questions will have to await the development of a rigorous general theory.

REFERENCES

1. Chamber, L. *The Geometrical Interpretation of Rayleigh and Schwinger Variational Principles. Proc. Int. Nat. Congress of Mathematicians*, vol. II, Amsterdam, 1954, p. 331.
2. Courant, R., and Hilbert, D., *Methoden der Mathematischen Physik*, vol. II; New York, Interscience Publishers, 1943.
3. Dolph, C., McLaughlin, J., and Marx, I. "Symmetric Linear Transformations and Complex Quadratic Forms." *Communications on Pure and Applied Mathematics*, vol. 7, 1954, pp. 621.
4. Dunford, N., "Spectral Operators," *Pacific Journal of Mathematics*, vol. 4, (1954), pp. 321.
5. Friedman, B., and Mishoe, L., *Eigenfunction Expansions Associated with a Non Self-Adjoint Differential Equation*. Res. Rept. B4-4, New York University, 1954.
6. Friedrichs, K., "Spektraltheorie halbbeschränkter Operatoren." *Mathematics Annual*, vol. 109 (1934), p. 465; *Mathematics Annual*, vol. 110 (1935), p. 777.
7. Hartree *et al.*, *Metorological Factors in Radio Wave Propagation*. Royal Meterological Socity, London, 1946.
8. Kato, T., "Notes on Schwinger Variational Method." *Progress of Theoretical Physics*, vol. 6, (1951), p. 295.
9. Kerr, D., *Propagation of Short Radio Waves*. McGraw-Hill Book Company, New York, 1951.
10. Kramer, H., *Perturbations of Differential Operators*. Berkeley Dissertation, No. 1954, Tech. Rept. No. 5, Cont. DA-040200-ORD-171.
11. Levine, H., *Variational Methods for Solving Electromagnetic Boundary Value Problem*. Lectures Published by Electronic Defense Laboratory, Mountain View, California, 1954.
12. Louhivarra, I., "Über das erste Randwertproblem für die Differentialgleichung $u_{xx} + u_{yy} + qu + f = 0$," *Annals of the Academy of Science of Finland*—Math. Phys. Ser. A, No. 183, 1955.
13. MacFarlane, G., "Variational Method for Determining Eigenvalues of the Wave Equation of Anomalous Propagation." *Proceedings of the Cambridge Philosophical Society*, vol. 43, (1947), p. 213.

14. Marcuvitz, N., *Field Representations in Spherically Stratified Regions*, Reprinted in *The Theory of Electromagnetic Waves—A Symposium*, New York, Interscience Publishers, 1951.

15. Marcuvitz, N., *Seminar in Recent Developments in the Theory of Wave Propagation*. Part III–D, New York University Notes, 1949–1950.

16. Morse, P., and Feshbach, H., *Methods of Theoretical Physics*, New York, McGraw-Hill Book Company, New York, 1954.

17. Müller, C., "Zur Methode der Strahlungskapazität von H. Weyl," *Mathematische Zeitung*, vol. 56, (1952), pp. 80–.

18. Naimark, M., *Investigation of the Spectra and Expansion Theorems of Non Self-adjoint Differential Operators on a Semi-axis*. Trudy Moscou, 3, 1954, p. 181–270 (Russian).

19. Nevanlinna, R., *Über Metrische Lineare Raüme*, Ann. of Acad. of Finland-Math-Physics, Ser. A, Nos. 108, 113, 115, (1952); No. 163, 1954.

20. Pekeris, C., Ament, S., *Characteristic Values of the First Normal Mode in the Problem of Propagation of Microwaves through Refraction*, Phil. Mag. Ser. 7, 38, 1947, p. 801.

21. Phillips, R. S., *Linear Ordinary Operators of the Second Order*, Research Rep. No. EM-42, Cont. AF–19(122)–42, New York University, 1952.

22. Riesz, F., St. Nagy, B., *Lecons D'analyse Fonctionelle*. Akademian Kiado, Budapest, 1952.

23. Saxon, D., "*Notes on Lectures by Julian Schwinger, Discontinuities in Wave Guides*, Massachusetts Institute of Technology Radiation Laboratory Series, 1945.

24. Schwartz, J., "Perturbations of Spectral Operators and Applications. I. Bounded Perturbations," *Pacific Journal of Mathematics*, vol. 4, (1954), p. 415.

25. Schwinger, J., and Levine, H., "On the Theory of Diffraction by an Aperture in an Infinite Plane Screen," I, *Physical Review*, vol. 74, (1948), pp. 958; II, *Physical Reviews*, vol. 75, (1946), p. 1423.

26. Schwinger, J., and Levine, H., *On the Theory of Electromagnetic Diffraction by an Aperture in an Infinite Plane Conducting Screen*, Reprinted in *The Theory of Electromagnetic Wave,—A Symposium*, New York, Interscience Publishers, 1952.

27. Sims, A., *Linear Differential Operators of the Second Order*, Dissertation, University of Southern California, June, 1954.

28. Sommerfeld, A., *Partial Differential Equations in Physics*, New York, Academic Press, 1947.

29. Titchmarsh, E., *Eigenfunction Expansions Associated with Second-Order Differential Equations*, Oxford, Oxford University Press, 1946.

30. von Karman, T., "The Engineer Grapples with non-linear Problems." *Bulletin of the American Mathematical Society*, vol. 46, (1940), p. 615.

31. Weyl, H., Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen." *Math. Ann.*, vol. 68, (1910), p. 220.

Diffraction by a Convex Cylinder*

JOSEPH B. KELLER†

Summary—The leading term in the asymptotic expansion for large k = $2\pi/\lambda$, of the fields reflected and diffracted by any convex cylinder are constructed. The cross section of the cylinder is assumed to be a smooth curve which may be either closed or open and extending to infinity. The method employed is an extension of geometrical optics in two respects. First, diffracted rays are introduced. Secondly, fields are associated with the rays in a simple way. The results are applicable when the wavelength is small compared to the cylinder dimensions.

I. INTRODUCTION

AS AN APPLICATION of a general method which has been devised for the asymptotic solution of diffraction problems,[1] we will treat the two-dimensional problem of diffraction of an arbitrary wave by a cylinder. For simplicity we will first consider an incident cylindrical wave. The field u will be assumed to be a scalar satisfying the reduced wave equation $(\nabla^2 + k^2)u = 0$ and vanishing on the cylinder. Later we will consider the vanishing of $\partial u/\partial \nu$ on the cylinder, and also the impedance boundary condition. Finally we will treat an arbitrary incident wave. The cross section of the cylinder will be assumed to be a smooth convex curve. If it extends to infinity (like a parabola) we will call it open; otherwise we will call it closed. For the case $u = 0$ on the cylinder we may interpret u as a component of electric field parallel to the cylinder generators, the cylinder as a perfect conductor, and the surrounding space as a homogeneous medium. The asymptotic solution to be obtained is the asymptotic expansion of u with respect to $k = 2\pi/\lambda$ as k tends to infinity (i.e., as the wavelength λ tends to zero and the frequency tends to infinity). The first term in this solution is just that obtained from geometrical optics. Thus the solution may be thought of as an improvement upon geometrical optics.

The method to be employed involves several simple steps. First the rays—incident, reflected and diffracted, are determined in Section II. Then the field associated with each ray is computed by essentially geometrical methods in Section III. In Section IV the results are assembled.

When applied to a circular cylinder in Section V the result agrees exactly with that obtained by Franz[2] and Imai[3] by asymptotic expansion of the exact solution.

In fact this comparison is used to determine a certain factor in the present method. In Section VI the result is applied to diffraction by a thick screen and by a wedge terminated by part of a circular cylinder. In Section VII diffraction by a parabolic cylinder is considered and exact agreement is found with the results of Rice[4] which are based upon asymptotic expansion of the exact solution. In Section IX the total cross section of the cylinder and related matters are discussed.

The method of this paper is closely related to the procedure used by Franz and Depperman[5] in their approximate treatment of diffraction by a circular cylinder.

II. RAY TRACING

Incident Rays

Since the medium is homogeneous, the incident rays are straight lines emanating from the source P of the incident cylindrical wave. Since no rays can penetrate the cylinder, there is a region called the geometrical shadow, which is devoid of incident rays.

Reflected Rays

Each incident ray which hits the cylinder gives rise to a reflected ray according to the law of reflection. Because the cylinder is smooth and convex, exactly one reflected ray will pass through each point not in the shadow. No reflected ray will reach any point in the shadow. Thus one incident and one reflected ray pass through each point outside the shadow. These two rays coincide for points on the shadow boundary. Neither an incident nor a reflected ray passes through any point in the shadow.

Diffracted Rays

The general definition of, and certain results pertaining to, diffracted rays have been given elsewhere.[1] Therefore we will describe here only those special aspects which are pertinent to the problem under consideration. The diffracted rays which occur in this problem may be characterized by the following extension of Fermat's Principle: The diffracted rays connecting the points P and Q are those curves which render stationary the Fermat integral (i.e., the optical length, or here, ordinary length) among all curves connecting P and Q and *having an arc on the cylinder*. It is easily seen that these rays consist of the tangents from P and Q to the cylinder plus that portion of the cylinder which connects the points of tangency (see Fig. 1). If the cylinder is closed there are two portions connecting these points of tangency. Only that portion which is a smooth con-

† Inst. of Math. Sci., New York Univ., New York, N. Y.

[1] J. B. Keller, "The Geometrical Theory of Diffraction," Proc. Symposium on Microwave Optics, McGill Univ., Montreal, Que., Can.; June, 1953.

[2] W. Franz, "Über die Greenschen funktionen des zylinders und der kugel," *Zeit. fur Naturf.*, vol. 9a, pp. 705–716; 1954.

[3] I. Imai, "Die beugung elektromagnetischer wellen an einen kreiszylinder," *Zeit. fur Phys.*, vol. 137, pp. 31–48; 1954.

[4] S. O. Rice, "Diffraction of a plane radio wave by a parabolic cylinder," *Bell Sys. Tech. Jour.*, pp. 417–504; March, 1954.

[5] W. Franz and K. Depperman, *Ann. der Phys.*, vol. 10, p. 361; 1952.

tinuation of the tangents yields a ray. However, in addition to this diffracted ray there are infinitely many others connecting P and Q. Each of these differs from the above merely in encircling the cylinder a finite number of times.
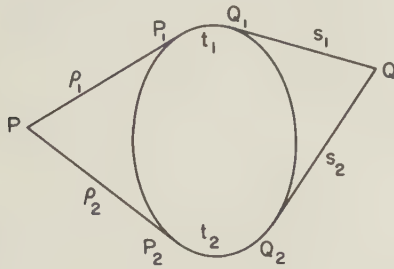
Fig. 1—The two rays diffracted from P to Q around a closed cylinder. The upper ray consists of a straight segment of length ρ_1 from P to the point of tangency P_1, an arc of length t from P_1 to Q_1 and a straight segment from Q_1 to Q, tangent at Q_1. The lower ray is similar.

Applying the preceding results to the present problem, we see that each incident ray which is tangent to the cylinder gives rise to a diffracted ray. There are either one or two such incident rays if the cross section extends to infinity, and exactly two if the cross section is closed. These are the grazing rays which bound the shadow. Each of the diffracted rays produced by these incident rays travels along the shaded surface of the cylinder. From each point of each of these surface rays a diffracted ray splits off along the tangent to the cylinder. Thus the shadow is covered by diffracted rays,—singly if the cylinder cross section is open and doubly if it is closed. The diffracted rays through each point Q are the (one or two) tangents from Q to the cylinder. In the case of cylinders of closed cross section, each point outside the shadow is also reached by two diffracted rays, since the surface ray continues around the cylinder into the lit region, then into the shadow again, etc. (see Fig. 2).

Fig. 2—The set of diffracted rays produced by a single incident ray. The incident ray is tangent to the cylinder and the diffracted rays are also tangent to it.

For such cylinders each diffracted ray actually consists of an infinite number of diffracted rays which have encircled the cylinder different numbers of times. The diffracted rays account for the field in the shadow

region. In the case of closed cylinders they also account for certain diffraction effects in the lit region.

III. FIELD CALCULATION

Field Associated With the Incident Rays

The optical form of the principle of conservation of energy states that in a steady-state (i.e., time-periodic) field the flux of energy is the same at all points of a narrow tube of rays. Furthermore this flux is proportional to the square of the field amplitude multiplied by the cross-sectional area of the tube and by the velocity of light at the point. Now in a cylindrical wave the distance between neighboring rays is $\rho d\theta$ where ρ is the distance from the source P and $d\theta$ is the angle between the two rays. In a homogeneous medium the velocity is constant. It follows that for a cylindrical wave in a homogeneous medium the amplitude A satisfies $A^2 \rho d\theta = A_o^2 d\theta$. Here A_o denotes the amplitude on the ray in question at unit distance from the source. From this equation we obtain $A_o/\sqrt{\rho}$ for the amplitude of a cylindrical wave at a distance ρ from the source P.

To determine the phase we note that in optics the difference between the phases at two points on a ray is just ω/c_o multiplied by the optical distance between the points. Here ω is the angular frequency of the field and c_o is the velocity of light in vacuum. If ρ is the distance between the points and c is the constant velocity of light in the medium, then the optical distance is just $c_o \rho/c$ and the phase is $\omega \rho/c = k\rho = 2\pi \rho/\lambda$ where $k = \omega/c = 2\pi/\lambda$. Thus knowing the phase at one point on a ray, we can find it at any other point on the same ray. If we assume that the phase is zero at the source $\rho = 0$, we find that the phase at a point with coordinate ρ is $k\rho$. Thus the field associated with an incident ray at a point with co-ordinate ρ is,

$$u_{\text{inc}} \sim \frac{A_o}{\sqrt{\rho}} \, e_{ik\rho} \,. \tag{1}$$

The field (1) is not a solution of the reduced wave equation. Instead it is the first term in the asymptotic expansion with respect to k, of such a solution. Additional terms can be found by the method of[6], but will not be considered here.

Field Associated With the Reflected Rays

In order to determine the field associated with a reflected ray, it is necessary to know the field on the incident ray and the conditions to be satisfied at the reflecting surface. We have already constructed the incident field. We must now impose the boundary condition that the field vanish at the surface. Then at the surface, the reflected field is the negative of the incident field. Therefore, at the surface the amplitudes of the reflected and incident fields are the same. However, the phase of the reflected field differs from that of the incident field by π. Thus the phase of the reflected field at a distance s from the surface along the reflected ray is

just the phase of the incident wave at the surface $(k\rho')$ plus π plus ks, i.e., $(k\rho' + ks + \pi)$. Here ρ' denotes the co-ordinate of the point of reflection.

To find the amplitude of the reflected field at a distance s from the surface along a reflected ray, we must consider a narrow tube of reflected rays containing the ray under consideration. In the present two-dimensional problem, it suffices to consider a pair of neighboring reflected rays. It is clear that two such reflected rays generally intersect if extended backward into the cylinder (see Fig. 3).
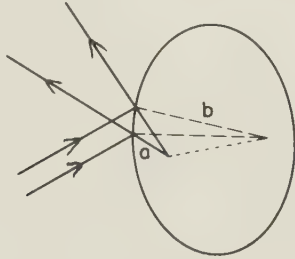
Fig. 3.—The reflected rays produced by a pair of parallel rays incident at angle θ on a cylinder. The reflected rays intersect at the distance $a = b/2 \cos \theta$ from the point of reflection, where b is the radius of curvature of the cylinder at that point.

Let us denote by a the distance from the point of reflection to the point of intersection. Then if $d\theta$ denotes the angle between the two rays, the cross-sectional area of the tube they bound (i.e., the distance between the rays) is $ad\theta$ at the reflecting surface and $(a + s)d\theta$ at the distance s from the surface along a ray. If $A(s)$ denotes the amplitude of the reflected field at s then the energy principle employed above yields

$$cA^2(s)(a + s)d\theta = cA_o{}^2ad\theta/\rho'.$$

From this we find $A(s)$, which is given by

$$A(s) = \frac{A_o}{\sqrt{\rho'}}(1 + a^{-1}s)^{-1/2}. \qquad (2)$$

In (2) ρ' is the distance from the source to the point of reflection. In case the neighboring reflected rays are parallel, (2) applies with a^{-1} replaced by zero (i.e., the rays intersect at infinity) and $A(s)$ remains constant along the reflected ray as in a plane wave.

Eq. (2) shows that in the present two dimensional problem the amplitude $A(s)$ diminishes inversely as the square root of the distance from the point of intersection of neighboring rays. Such an intersection point is called a caustic point, and the locus of such points is called the caustic of the ray system. The amplitude given by (2) is infinite at a caustic. However, the caustic does not occur on the real parts of the rays, but instead on their fictitious backward extensions. Therefore, in the present problem no difficulty due to a caustic arises. Of course the singularity of (2) implies that the energy principle of geometrical optics, from which (2) was deduced, is not valid at a caustic.

Before (2) can be used, the distance a must be determined. To find a we note that neighboring normals to the reflecting surface meet at a distance b from the

surface, and b is called the radius of curvature of the surface at the point under consideration. For an incident plane wave any two neighboring incident rays are parallel. If they make the angle ϕ with the two neighboring normals, then $a = \frac{1}{2}b \cos \phi$ (see Fig. 3). This value of a is the focal length of the reflector for the particular point and angle of incidence considered. Now for an incident cylindrical wave we have from the mirror law of optics

$$\frac{1}{a} = \frac{2}{b \cos \phi} + \frac{1}{\rho'}. \qquad (3)$$

The reflected field can now be obtained by combining our result for the phase with (2) and using the above value of $1/a$. Then we obtain

$$u_{\text{ref}} \sim -\frac{A_o}{\sqrt{\rho'}}\left[1 + \frac{2s}{b \cos \phi} + \frac{s}{\rho'}\right]^{-1/2}e^{ik(\rho'+s)}. \qquad (4)$$

The reflected field (4) which has been computed by the principles of geometrical optics, is not a solution of the reduced wave equation. Therefore it is not a complete representation of the reflected field. We expect it to be the first term in the asymptotic expansion of the reflected field with respect to k. Additional terms have been determined,[6] but will not be required here.

Field Associated With the Diffracted Rays

Consider a diffracted ray which originates at a point of tangency P_1 and leaves the cylinder tangentially at another point Q_1 (see Fig. 1). Let ρ_1 be the co-ordinate of P_1, and let t be the length of the ray from P_1 to Q_1. Then the phase of the incident field is $k\rho_1$ at P_1 and $k(\rho_1 + t)$ at Q_1. At a point a distance s along the diffracted ray from Q_1 the phase is $k(\rho_1 + t + s)$.

If the cylinder is closed, there are two points, P_1 and P_2, at which diffracted rays originate. If the co-ordinate of P_2 is ρ_2 then the phase of the diffracted ray originating at P_2 and leaving the surface at Q_2 is $k(\rho_2 + t + s)$. Here t denotes the length of the ray from P_2 to Q_2. For the closed cylinder there are infinitely many values of t corresponding to the same pair of points P_1 and Q_1 (or P_2 and Q_2). These values differ from each other by multiples of T, the length of the cross-sectional curve of the cylinder. They correspond to rays which encircle the cylinder a finite number of times.

Let us denote by $A(s, t, P_1, Q_1)$ the amplitude of the diffracted field on a ray originating at P_1 and leaving the cylinder at Q_1. First we will consider the dependence of A on s, the distance along the diffracted ray from the surface. To this end we must consider the cross-sectional area of a tube of neighboring diffracted rays. In the present two-dimensional case this is just the distance between two neighboring diffracted rays. Now two such rays interesect on the cylinder since they are tangent to it. Therefore the distance between two such rays is

[6] J. B. Keller, R. M. Lewis, and B. Seckler, "Asymptotic Solution of Some Diffraction Problems," Res. Rep. EM–81, Inst. Math. Sci., New York Univ., New York, N. Y.; June, 1955.

$sd\theta$ where $d\theta$ is the angle between them. By the energy principle applied to this tube,

$$A^2(s, t, P_1, Q_1)sd\theta = A^2(1, t, P_1, Q_1) d\theta \qquad (5)$$

If we set $A(1, t, P_1, Q_1) = a(t, P_1, Q_1)$ then this yields

$$A(s, t, P_1, Q_1) = \frac{a(t, P_1, Q_1)}{\sqrt{s}} \qquad (6)$$

From (6) we see that A decreases like $s^{-1/2}$. However, on the cylinder, where $s = 0$, A becomes infinite. Thus (6) does not apply on the cylinder, which is a caustic of the diffracted field.

In order to determine $a(t, P_1, Q_1)$ we observe that at each of its points the diffracted ray sheds a ray along the tangent to the surface. This tangent ray must take some energy away from the diffracted ray on the surface. This energy must be proportional to the energy of the surface ray at the point of tangency. The proportionality factor may depend upon the local properties of the surface. Therefore, we assume

$$\frac{da(t, P_1, Q_1)}{dt} = -\alpha(t)a(t, P_1, Q_1). \qquad (7)$$

Here $\alpha(t)$ is the proportionality factor. From (7) we find

$$a(t, P_1, Q_1) = a(0, P_1, Q_1) \exp\left[-\int_o^t \alpha(\tau) \, d\tau\right]. \qquad (8)$$

To determine the coefficient $a(0, P_1, Q_1)$ we first note that it must be proportional to the amplitude of the incident field at P_1 since this field produces the diffracted ray. Apart from this dependence, $a(0, P_1, Q_1)$ must be symmetric in P_1 and Q_1. This symmetry is a consequence of the reciprocity theorem for u which states that the field produced at Q by a source at P is the same as the field produced at P by a source at Q. When applied to the asymptotic expansion of u, which we are constructing, this theorem implies the symmetry referred to above. Finally, we note that the diffraction processes at P_1 and Q_1 are independent of each other. This means that any changes in the cylinder shape at P_1 will produce corresponding changes in the field whatever the shape of the cylinder at Q_1. Consequently $a(0, P_1, Q_1)$ must be a product of a function of P_1 and a function of Q_1. Thus we have

$$a(0, P_1, Q_1) = \frac{A_o}{\sqrt{\rho_1}} B(P_1)B(Q_1). \qquad (9)$$

The function $B(P_1)$ or $B(Q_1)$ is independent of the nature of the incident wave, in the small wavelength limit with which we are dealing. This is so because only that part of the wave near P_1 or Q_1 plays a role, and every wave is locally plane.

Collecting our results we have for the field diffracted by a cylinder with an open cross section

$$u_{\text{diff}} \sim \frac{A_o B(P_1)B(Q_1)}{\sqrt{\rho_1 s}} \exp\left[-\int_o^t \alpha(\tau) \, d\tau + ik(\rho_1 + t + s)\right]. \qquad (10)$$

If the cylinder has a closed cross section then t takes on the values $t + mT$, $m = 0, 1, 2, \cdots$. Therefore u_{diff} is the sum of terms of the form (10), one for each value of m. This sum is just a geometric series, which can be summed at once. In addition there is another such sum corresponding to the rays originating at P_2. Thus, for a closed cylinder

$$u_{\text{diff}} \sim \frac{A_o B(P_1)B(Q_1)}{\sqrt{\rho_1 s_1}}$$
$$\exp\left[-\int_o^{t_1} \alpha(\tau) \, d\tau + ik(\rho_1 + t_1 + s_1)\right]$$
$$\cdot \left\{1 - \exp\left[-\int_o^T \alpha(\tau) \, d\tau + ikT\right]\right\}^{-1}$$
$$+ \frac{A_o B(P_2)B(Q_2)}{\sqrt{\rho_2 s_2}} \exp\left[-\int_o^{t_1} \alpha(\tau) \, d\tau + ik(\rho_2 + t_2 + s_2)\right]$$
$$\cdot \left\{1 - \exp\left[-\int_o^T \alpha d\tau + ikT\right]\right\}^{-1}. \qquad (11)$$

In (11) t_1 denotes distance along a ray from P_1 to Q_1, and s_1 distance along a tangent ray from Q_1. Corresponding definitions apply to t_2 and s_2.

In order to complete the determination of u_{diff} we must determine $B(P_1)$ and $\alpha(t)$. These quantities are related to the field on and near the caustic (i.e., the cylinder surface) where the expansions (10) and (11) fail. Therefore, we must make use of some other form of solution valid on the caustic in order to determine these quantities. However this solution need not be a solution of the present problem. It need only have the same kind of caustic as does the present problem. Thus to determine $\alpha(t)$ we require a solution which represents a wave traveling around a curved cylinder. For this purpose an exact solution describing a wave traveling around a circular cylinder of radius b is obtained in Appendix I. Then (10) is applied to the same cylinder. The two results are found to agree provided $\alpha = \alpha(b)$ has the value (12). We assume that α is the same function of $b(t)$ when b varies.

The result is that there are infinitely many values of α given by

$$\alpha_n = -ik^{1/3}b^{-2/3}\tau_n \qquad (12)$$

where

$$\tau_n = \tfrac{1}{2}[3\pi(n + \tfrac{3}{4})]^{2/3}e^{i\pi/3}, \quad n = 0, 1, 2, \cdots. \qquad (13)$$

The diffracted fields corresponding to these different values of α may be called the modes of the cylinder. Since various modes are possible, we must replace (10) and (11) by sums over the various modes. The coefficient $B(P_1)$ must then depend upon the mode, so we will denote the coefficients by $B_n(P_1)$. Now we replace (10) and (11), which give the fields diffracted respectively by an open and a closed cylinder hit by a cylindrical wave, by

$$u_{\text{diff}} \sim \frac{A_o e^{ik(\rho_1 + t + s)}}{\sqrt{\rho_1 s}} \sum_{n=o}^{\infty} B_n(P_1)B_n(Q_1)$$
$$\exp\left[i\tau_n k^{1/3}\int_o^t b^{-2/3}(\tau) \, d\tau\right] \qquad (10a)$$

$$u_{\text{diff}} \sim \frac{A_o e^{ik(\rho_1 + t_1 + s_1)}}{\sqrt{\rho_1 s_1}} \sum_{n=o}^{\infty} B_n(P_1) B_n(Q_1)$$

$$\exp\left[i\tau_n k^{1/3} \int_o^{t_1} b^{-2/3}(\tau) \, d\tau\right]$$

$$\left\{1 - \exp\left[ikT + i\tau_n k^{1/3} \int_o^T b^{-2/3}(\tau) \, d\tau\right]\right\}^{-1}$$

$$+ \frac{A_o e^{ik(\rho_1 + t_2 + s_2)}}{\sqrt{\rho_2 s_2}} \sum_{n=o}^{\infty} B(P_2) B(Q_2)$$

$$\exp\left[i\tau_n k^{1/3} \int_o^{t_2} b^{-2/3}(\tau) \, d\tau\right]$$

$$\left\{1 - \exp\left[ikT + i\tau_n k^{1/3} \int_o^T b^{-2/3}(\tau) \, d\tau\right]\right\}^{-1}. \quad (11a)$$

These results must now be completed by the determination of $B_n(P_1)$. As we have seen $\alpha_n(t)$ is independent of where or how the diffracted ray begins. $B_n(P_1)$, on the other hand, depends essentially on the way in which the diffracted wave is produced by the incident wave. Therefore to determine it we must consider the solution of a problem in which a wave is incident upon some cylinder and produces a diffracted wave. From this problem we will determine $B_n(P_1)$ which will depend upon the local geometry (i.e., curvature) of the cylinder at P_1. For this purpose we will make use of the solution corresponding to a plane wave incident on a circular cylinder. We will assume, of course, that the dependence of $B_n(P_1)$ on b_1, the curvature at P_1, is just that determined from the solution for a circular cylinder, i.e., $B_n(P_1) = B_n(b_1)$.

We could determine both $B_n(P_1)$ and $\alpha_n(b)$ from the solution corresponding to a plane wave incident upon a circular cylinder. However, it seemed preferable to first determine $\alpha_n(b)$ and then to determine $B_n(b_1)$ separately in the manner indicated above. In this way the physical significance of the two quantities is made more apparent. Furthermore, the determination of $\alpha_n(b)$ was based on a relatively simple solution of the wave equation, while that of $B_n(b_1)$ depends upon a more complicated solution.

In Appendix II, (11a) is applied to a circular cylinder. The result is found to coincide exactly with that of Franz[7], which is obtained by asymptotic expansion of the explicit solution of this problem, provided that $B_n(b)$ is given by

$$B_n(b) = b^{1/6} k^{-1/12} e^{i\pi/24} \left[\frac{\pi\sqrt{8\pi}}{12(6^{2/3})[A'(6^{1/3} \mid \tau_n \mid)]^2}\right]^{1/2}. \quad (14)$$

Here $A'(6^{1/3} \mid \tau_n \mid)$ is the derivative of the Airy integral

$$A(6^{1/3} \mid \tau_n \mid) = \int_o^{\infty} \cos(z^3 - 6^{1/3} \mid \tau_n \mid z) \, dz. \quad (15)$$

---

[7] Franz, *op. cit.*, eq. (25a).

IV. The Total Field

With the determination of B_n, the calculation of the field is complete. We may now summarize our results by writing

$$u = u_{\text{inc}} + u_{\text{ref}} + u_{\text{diff}}. \quad (16)$$

$$u_{\text{inc}} \sim \frac{A_o}{\sqrt{\rho}} e^{ik\rho} \quad (17)$$

$$u_{\text{ref}} \sim -A_o\left(\rho' + \frac{2\rho's'}{b\cos\phi} + s'\right)^{-1/2} e^{ik(\rho'+s')}. \quad (18)$$

For an open cylinder

$$u_{\text{diff}} \sim \frac{A_o}{\sqrt{2ks\rho_1}} k^{1/3}[b(P_1)b(Q_1)]^{1/6} e^{ik(\rho_1 + t + s) + i\pi/12}$$

$$\sum_{n=o}^{\infty} C_n \exp\left[i\tau_n k^{1/3} \int_o^t b^{-2/3}(\tau) \, d\tau\right]. \quad (19)$$

For a closed cylinder

$$u_{\text{diff}} \sim \frac{A_o e^{i\pi/12}}{\sqrt{2ks_1\rho_1}} k^{1/3}[b(P_1)b(Q_1)]^{1/6} e^{ik(\rho_1 + t_1 + s_1)} \sum_{n=o}^{\infty} C_n$$

$$\exp\left[i\tau_n k^{1/3} \int_o^{t_1} b^{-2/3}(\tau) \, d\tau\right]$$

$$\left\{1 - \exp\left[ikT + i\tau_n k^{1/3} \int_o^T b^{-2/3}(\tau) \, d\tau\right]\right\}^{-1}$$

$$+ \frac{A_o e^{i\pi/12}}{\sqrt{2ks_2\rho_2}} k^{1/3}[b(P_2)b(Q_2)]^{1/6} e^{ik(\rho_1 + t_2 + s_1)} \sum_{n=o}^{\infty} C_n$$

$$\exp\left[i\tau_n k^{1/3} \int_o^{t_2} b^{-2/3}(\tau) \, d\tau\right]$$

$$\left\{1 - \exp\left[ikT + i\tau_n k^{1/3} \int_o^T b^{-2/3}(\tau) \, d\tau\right]\right\}^{-1} \quad (20)$$

$$\tau_n = \tfrac{1}{2}[3\pi(n + \tfrac{3}{4})]^{2/3} e^{i\pi/3} \quad (21)$$

$$C_n = \pi^{3/2}\left\{3(6^{1/3})[A'(6^{1/3} \mid \tau_n \mid)]^2\right\}^{-1}. \quad (22)$$

For $n = 0, 1$ the values of C_n and more precise values of τ_n are [2,3]

$$\tau_o = 1.8557571 e^{i\pi/3}, \qquad \tau_1 = 3.2446076 e^{i\pi/3} \quad (23)$$

$$C_o = 0.9107193, \qquad C_1 = 0.6942728. \quad (24)$$

In (18) ρ', b and ϕ are the distance from the source to the point of specular reflection, the radius of curvature and the angle of incidence at this point respectively, while s' is distance from this point. In (19) and (20) P_1 and P_2 denote points of tangency at which the diffracted rays originate, $b(P_1)$ and $b(P_2)$ denote the radii of curvature at these points and ρ_1, ρ_2 denote the distances of these points from the source. Q_1 and Q_2 are the points at which the diffracted rays leave the surface, s_1 and s_2 are the distances from these points to the field point and $b(Q_1)$, $b(Q_2)$ are the radii of curvature at these points. The distances from P_1 to Q_1 and from P_2 to Q_2 along the cylinder are denoted by t_1 and t_2 respectively.

The results (19) and (20) become infinite on the cylinder since on it $s = 0$ (or $s_1 = s_2 = 0$). To obtain a finite value for the field on and near the cylinder, we again make use of the exact solution for a wave traveling

around a circular cylinder. In Appendix III we compare this solution with (19) specialized to a circular cylinder. This shows us how to modify (19) in order that it yield a finite field on and near the circular cylinder. Then we assume that this same modification applies for any cylinder. In this way we obtain for the field near an open cylinder, instead of (19),

$$u_{\text{diff}} \sim 2^{3/4} \frac{A_o}{\sqrt{\rho_1}} e^{ik(\rho_1 + t + s)} \sum_{n=0}^{\infty} B_n(P_1) B_n(Q_1)$$

$$\{[k^{-1}b^2(t)]^{2/3}\tau_n - s^{3/2}]^{-1/4}$$

$$\exp\{i\tau_n k^{1/3}[b^{-2/3}(t)s + \int_o^t b^{-2/3} dt]\}$$

$$\cos\left\{\frac{\pi}{4} - \frac{i}{3}(kb)^{1/2}[2(kb)^{1/3}\tau_n - ks^2 b^{-1}]^{3/2}\right\}. \quad (25)$$

The corresponding expression for a closed cylinder consists of two terms of the above form. However, each term in the summation is multiplied by

$$[1 - \exp(ikT + i\tau_n k^{1/3} \int_o^T b^{-2/3} d\tau)]^{-1}.$$

For $s^2 << |\tau_n|(k^{-1}b^2)^{2/3}$ we see that each term in (25) is proportional to k^o. On the other hand, each term in (19) is proportional to $k^{-1/6}$. Thus we see that the diffracted field near the surface is of the order $k^{1/6}$ greater than that away from the surface. This is exactly the behavior to be expected at a caustic.[8]

V. Diffraction by a Circular Cylinder

As an application of our results, let us determine u_{diff} for a circular cylinder by using (20). Let the source be at $(R, 0)$, the observation point at (r, θ) and let $A_o = \sqrt{i/8\pi k}$. Then utilizing the results of Appendices 1 and 2, we find that (20) becomes

$$u_{\text{diff}} \sim \sqrt{i/\pi} \frac{\exp\{ik[(R^2 - b^2)^{1/2} + (r^2 - b^2)^{1/2}] + i\pi/12\}}{4k(R^2 - b^2)^{1/4}(r^2 - b^2)^{1/4}}$$

$$(kb)^{1/3}$$

$$\sum_{n=0}^{\infty} C_n \frac{\exp\{i[kb + (kb)^{1/3}\tau_n]\theta\} + \exp\{i[kb + (kb)^{1/3}\tau_n](2\pi - \theta)\}}{1 - \exp[2\pi i[kb + (kb)^{1/3}\tau_n]]}$$

$$(26)$$

$$\exp\{i[kb + (k+b)^{1/3}\tau_n][\cos^{-1}(b/r) + \cos^{-1}(b/R)]\}.$$

The above choice of A_o corresponds to the incident field $i/4 H_o^{(1)}(k\rho)$.

It has already been pointed out that the preceding result for the field diffracted by a circular cylinder with a cylindrical incident wave agrees exactly with the asymptotic expansion of the known solution.[2] In (2) the reflected field was determined also, but only for an incident plane wave, and then only at great distances from the cylinder. Our result (18) coincides with the value given in (2) this case. Similar agreement is found with the results of Imai,[3] who treated a plane wave

[8] I. Kay and J. B. Keller, "Asymptotic evaluation of the field at a caustic," *Jour. Appl. Phys.*, vol. 25, pp. 876–883; July, 1954.

incident upon a circular cylinder. His results for both the reflected and diffracted fields are given only at great distances from the cylinder.

VI. Diffraction by a Thick Screen with a Circular End

As a second application of our result, let us consider the diffraction of a cylindrical wave by a thick screen with a circular end (see Fig. 4). The cross section of this screen
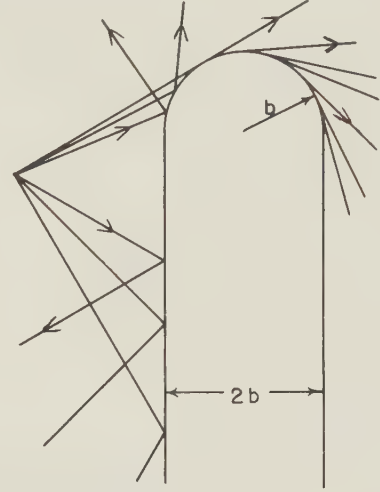
Fig. 4—Reflected and diffracted rays produced by the rays from a line source when they hit a thick screen with a circular end. The width of the screen is 2b and the radius of the semicircle is b.

consists of the two lines $x = \pm b$, $y < 0$ and the semicircle $r = b$, $y = 0$. Formulas (17), (18), and (19) immediately apply to this problem. The radius of curvature is b on the semicircle and infinite on the plane. Since the points of tangency P_1 and Q_1 are on the semicircle, $b(P_1) = b(Q_1) = b$ and in the integral in (19) $b(\tau) = b$. If we express t, ρ, and s in terms of the co-ordinates R, θ' of the source and r, θ of the observation point, then the diffracted field becomes

$$u_{\text{diff}} \sim \frac{A_o(kb)^{1/3} e^{ik[(R^2 - b^2)^{1/2} + (r^2 - b^2)^{1/2}] + [i\pi/12]}}{\sqrt{2k}(R^2 - b^2)^{1/4}(r^2 - b^2)^{1/4}} \quad (27)$$

$$\sum_{n=0}^{\infty} C_n e^{i[kb + (kb)^{1/3}\tau_n][\theta - \theta' - \cos^{-1}(b/r) - \cos^{-1}(b/R)]}.$$

The diffracted field occurs only at points through which diffracted rays pass. These points are those for which the inequality $\theta - \theta' - \cos^{-1}(b/r) - \cos^{-1}(b/R) > 0$ is satisfied.

The above result also applies if the screen is replaced by a wedge with a circular end. In this case the two lines may be any two tangents to the circle. Neither this problem, nor that of the screen, has been solved previously so we have no other results with which to compare ours.

It should be observed that the reflected field given by (18) is discontinuous on the ray reflected from the point where the circle and the straight side of the screen meet. This is due to the fact that the radius of curvature is

discontinuous at this point. An asymptotic expression different from that found here applies on this line. This is also the case on the shadow boundary, where the present result is invalid. The expansion on and near the shadow boundary in this problem has been considered by Artmann.[9]

VII. Diffraction by a Parabolic Cylinder

As another application we will consider the diffraction of a cylindrical wave by a parabolic cylinder. Let the cross section of the cylinder be the parabola $x^2 = 4h(h - y)$ (see Fig. 5) and let the source be to the left of $y = 0$. Again the results (17), (18) and (19) apply.
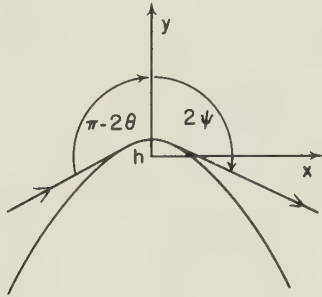
Fig. 5—A ray incident tangentially on a parabolic cylinder and one of the resulting diffracted rays.

The radius of curvature of the parabola is $b(x) = 2h[1 + (x/2h)^2]^{3/2}$. If x_1, and x denote the x-coordinates of the points of tangency P_1 and Q_1, then we have, assuming $x_1 < x$,

$$t = \int_{x_1}^{x} \left[1 + \left(\frac{x}{2h}\right)^2\right]^{1/2} dx$$

$$= \frac{x}{2}\left[1 + \left(\frac{x}{2h}\right)^2\right]^{1/2} - \frac{x_1}{2}\left[1 + \left(\frac{x_1}{2h}\right)^2\right]^{1/2}$$

$$+ h \log\left\{\frac{x}{2h} + \left[1 + \left(\frac{x}{2h}\right)^2\right]^{1/2}\right\}$$

$$\left\{\frac{x_1}{2h} + \left[1 + \left(\frac{x_1}{2h}\right)^2\right]^{1/2}\right\}^{-1} \quad (28)$$

$$\int_{x_1}^{x} b^{-2/3}\left[1 + \left(\frac{x}{2h}\right)^2\right]^{1/2} dx$$

$$= (2h)^{1/3} \log \frac{x + [x^2 + 4h^2]^{1/2}}{x_1 + [(x_1)^2 + 4h^2]^{1/2}}. \quad (29)$$

Before inserting these expressions into (19) we can simplify them by introducing the angle 2θ and 2ψ. These are respectively the angles (measured clockwise) between the positive y-axis and the incident and diffracted rays, i.e, the tangents to the parabola at P_1 and Q_1. From the tangency condition we have

$$\cot 2\theta = -\frac{x_1}{2h} \qquad \cot 2\psi = -\frac{x}{2h}. \quad (30)$$

[9] K. Artmann, "Beugung polarisierten lichtes an biegenden endlicher dicke im gebiet der Schattengrenze," *Zeit. fur Phys.*, vol. 127, pp. 408–494; 1950.

Now we find

$$\frac{x}{2h} + \left[1 + \left(\frac{x}{2h}\right)^2\right]^{1/2} = -\cot 2\psi + \csc 2\psi = \tan \psi.$$

$$(31)$$

A corresponding relation applies with x_1 and θ in place of x and ψ. Also the expression above for the radius of curvature, becomes

$$b(P_1) = 2h \csc^3 2\theta; \qquad b(Q_1) = 2h \csc^3 2\psi. \quad (32)$$

Now upon inserting the above expressions into (19) we obtain

$$u_{\text{diff}} \sim \frac{A_o}{\sqrt{2ks\rho_1}} (2kh)^{1/3}[\csc 2\theta \csc 2\psi]^{1/2}$$

$$\exp\left[ik\left(\rho_1 + s + \frac{x}{2}\csc 2\psi - \frac{x_1}{2}\csc 2\theta\right) + \frac{i\pi}{12}\right]$$

$$\sum_{n=0}^{\infty} C_n\left[\frac{\tan \psi}{\tan \theta}\right]^{ikh + i\tau_n(2kh)^{1/3}}. \quad (33)$$

This result (33) for the diffracted field is valid only in the region covered by diffracted rays. If the source is to the left of the y-axis, as we have assumed, this is the region $\psi > \theta$. The angle 2θ and the distance ρ_1 are constants determined by the location of the source. The angle ψ and distance s locate the point at which u_{diff} is evaluated.

Suppose that the polar co-ordinates of the source and observation points are respectively R, θ' and r, ψ' with the positive x-axis as polar axis. Now if $R >> h, r >> h$ then we see that $\theta' \sim 3\pi/2 - 2\theta, \psi' \sim \pi/2 - 2\psi, R \sim \rho_1 - x'/2 \csc 2\theta$ and $r \sim s + x/2 \csc 2\psi$. Then (33) becomes

$$u_{\text{diff}} \sim \frac{A_o}{\sqrt{2krR}} (2kh)^{1/3}[-\sec \theta' \sec \psi']^{1/2}$$

$$\exp\left[ik(R + r) + \frac{i\pi}{12}\right]$$

$$\cdot \sum_{n=0}^{\infty} C_n\left[-\tan\left(\frac{\psi'}{2} - \frac{\pi}{4}\right)\tan\left(\frac{\theta'}{2} - \frac{\pi}{4}\right)\right]^{[ikh + i\tau_n(2kh)^{1/3}]}. \quad (34)$$

The region covered by diffracted rays is now defined by $\psi' > \theta' - \pi$.

The expression (34) for u_{diff} can be compared with the result of Rice[4] provided that we specialize (34) to the case of an incident plane wave. To do this we omit $R^{1/2}$ from the denominator, replace R in the exponent by $x_1 = 2h \tan(\theta'/2 + \pi/4)$ and denote the incident wave by $e^{-ik(x \cos \theta' + y \sin \theta')}$. The plane wave is then propagating in the direction $\theta' + \pi$. The comparable result of Rice was obtained by asymptotic expansion of the exact solution, which is a series of products of parabolic cylinder functions. It is given in his equation (7.14).

We find that (34) and Rice's result agree exactly provided we correct what appears to be a slight error in his result. He gives an expression for the s-th zero of the parabolic cylinder functions which may be written in our notation as

$$n_s = i[kh + (2kh)^{1/3}\tau_s] - 1.$$

We find agreement if we replace the -1 by $-\frac{1}{2}$. This change makes Rice's expression for the zero agree with the results of Schwid[10] which are considered to be correct. Rice mentions the discrepancy between his and Schwid's results. We have found an error in Rice's analysis, which when corrected leads to the $-\frac{1}{2}$. Therefore we conclude that this change should be made in Rice's result and that then his and our results agree and are correct.

VIII. NORMAL DERIVATIVE AND IMPEDANCE BOUNDARY CONDITIONS

The preceding method applies directly to the corresponding problems in which either the normal derivative of u vanishes on the boundary or else u satisfies the impedance boundary condition $\partial u/\partial \nu = ikZu$. In the latter case the impedance Z is a given characteristic of the cylinder, which may vary with position on the surface. In both cases the rays are the same as in the present case $u = 0$ and the incident field is unchanged. The reflected field (4) is changed by the replacement of the minus sign by a plus sign when $\partial u/\partial \nu = 0$ since the phase is not changed by reflection in this case. In the impedance case the minus sign is replaced by the reflection coefficient $(\cos \phi - Z)/(\cos \phi + Z)$. This coefficient is obtained merely by inserting the expressions for the incident and reflected fields into the impedance condition and solving for the reflection coefficient. The normal is considered to point into the cylinder and ϕ is the angle of incidence.

In determining the diffracted field the only changes occur in the expressions for τ_n, B_n and C_n. For the normal derivative case (13) and (21) must be replaced by

$$\tau_n = \tfrac{1}{2}[3\pi(n + \tfrac{1}{4})]^{2/3}e^{i\pi/3}, \quad n = 0, 1, 2, \cdots. \quad (35)$$

The expression for B_n needn't be given since the final result contains only C_n. C_n is no longer given by (22) but instead by

$$C_n = \pi^{3/2}\{6^{2/3}\,|\,\tau_n\,|\,A^2(6^{1/3}|\,\tau_n\,|)\}^{-1}. \quad (36)$$

For $n = 0, 1$ the values of C_n and more precise values of τ_n are [2,3]

$$\tau_o = 0.8086166e^{i\pi/3} \qquad \tau_1 = 2.5780962e^{i\pi/3} \quad (37)$$

$$C_o = 1.5322785 \qquad C_1 = 0.7851980. \quad (38)$$

In the case of the impedance boundary condition the appropriate values of τ_n are determined by the equation

$$\frac{\partial}{\partial(kb)}H_\nu^{(1)}(kb) = iZH_\nu^{(1)}(kb), \quad (39)$$

where $\nu = kb + (kb)^{1/3}\tau$. Upon using the expansion $(A4)$ for $H_\nu^{(1)}(kb)$ this equation becomes

$$(2\tau)^{1/2}\tan\left[\frac{\pi}{4} - \frac{i}{3}(2\tau)^{3/2}\right] = -(kb)^{1/3}Z. \quad (40)$$

The roots τ_n of this equation are discussed by Bremmer[11].

It is to be noted that since τ_n depends upon both b and Z it will generally vary with position on the cylinder. Therefore it cannot be taken outside the integral signs as in (19) and (20).

The corresponding coefficients B_n and C_n can also be found by comparing the present solution with the asymptotic expansion of the exact solution for a circular cylinder. The impedance in the circular cylinder problem may be taken as constant and equal to the value at the point of initiation of the diffracted ray. Upon performing the necessary calculations we find

$$C_n = \pi^{3/2}[6^{2/3}\,|\,\tau_n\,|\,A^2 + 3e^{i5\pi/6}\,(kb)^{1/3}Z\,A\,A']^{-1}. \quad (41)$$

These considerations complete the solution for the impedance boundary condition. The Airy function in (41) is evaluated at $6^{1/3}\,|\,\tau_n\,|$.

IX. CONCLUSION

We have given a method for constructing the leading terms in the asymptotic expansion of the fields reflected and diffracted by a smooth convex cylinder. The formulas were derived for an incident cylindrical wave. However, the method applies equally well to any incident wave. With minor modifications the results do too. Thus, for example consider formula (19) for the diffracted field. The factor $A_o e^{ik\rho_1}/\sqrt{\rho_1}$ represents the cylindrical wave at the point of initiation of the diffracted ray. We need merely replace this by the value of the arbitrary incident wave at the corresponding point and (19) remains valid. A similar change applies to (20). Eq. (18) for the reflected field remains valid if ρ' in the phase is replaced by the phase of the incident wave at the point of reflection. The ρ' in the amplitude is the radius of curvature of the incident wavefront at the same point.

The results of this method have been checked against asymptotic expansions of exact solutions for the circular and parabolic cylinders. Exact agreement has been found in the case of the boundary condition $u = 0$, mentioned previously, and also in the case of the boundary condition $\partial u/\partial \nu = 0$. In the case of the impedance boundary condition agreement has been found only with the solution for the circular cylinder, since the parabolic cylinder problem has not been solved explicitly.

Additional terms in the expansion of the reflected field can be computed by the method mentioned.[6] Further terms in the expansion of the diffracted field can be obtained by using the method of Friedlander and the author.[12] However, no procedure is yet available for determining the initial conditions required in that method.

Some of the results of the present paper have also been verified in a different way by I. Kay and the author. We applied to an arbitrary cylinder the integral equation method used by Franz and Depperman[5] for a

[10] N. Schwid, "The asymptotic forms of the Hermite and Weber functions," *Amer. Math. Soc. Trans.*, vol. 37, pp. 339–362; 1935.

[11] H. Bremmer, "Terrestrial Radio Waves," Elsevier, New York, N. Y.—Houston, Tex., p. 39, eq. (14a); 1949.

[12] F. G. Friedlander and J. B. Keller, "Asymptotic expansions of solutions of $(\nabla^2 + k^2)u = 0$, *Comm. Pure and Appl. Math.*, vol. 8, pp. 382–394; 1955.

circular cylinder. In this way we again obtained the same value of $\alpha(t)$ as that given here. We also found that there is no diffracted field of the type considered here on a concave surface. This analysis is omitted since it is so similar to that of Franz and Depperman and also because the method of the present paper is much simpler.

The asymptotic expansion of the field within a few wavelengths of the cylinder is given by a different expression from that valid further from the cylinder. In the same way, a different asymptotic expansion is required near the shadow boundary. It can be obtained from the results of Rice[4] or of Artmann.[9] However, since neither of their results is exact, the expansion valid near the shadow boundary will not be given here. An investigation of this region based upon the exact solution for the circular cylinder is nearly completed. When that investigation is complete such an expansion can be given.

The far field in the forward scattering direction is required for the determination of the total scattering cross section of the cylinder, and this direction is parallel to that of the shadow boundary. Therefore the total cross section cannot be determined until the expansion valid near the shadow boundary is known for an incident plane wave. By making use of the results of Rice, of Artmann, or the as yet incomplete calculation just referred to, we find

$$\frac{\sigma}{2g} = 1 + \frac{M}{k^{2/3}g}(b_1{}^{1/3} + b_2{}^{1/3}). \quad (42)$$

Here σ is the scattering cross section of the cylinder (per unit length) and $2g$ is the geometrical optics cross section, i.e., g is the width of the cylinder normal to the direction of propagation of the plane wave. The radii of curvature at the two points of tangency of the incident rays are b_1 and b_2 and M is a constant which depends upon the boundary condition.

From recent work of G. Kear on the circular cylinder we find for the constant M

$$M = +0.48309 \qquad u = 0 \quad (43)$$

$$M = -0.432 \qquad \frac{\partial u}{\partial \nu} = 0. \quad (44)$$

Applying (42) to the circular cylinder yields

$$\frac{\sigma}{4b} = 1 + M(kb)^{-2/3}.$$

It should be noted that the above result for the total cross section does not oscillate as k varies. Even the more complete expression for σ which contains the exponential terms of the diffracted field does not oscillate until kb is nearly equal to unity. This is due to the fact that the diffracted rays must encircle the cylinder at least once before proceeding in the forward direction. In encircling the cylinder the field on these rays decays exponentially to such a small value that it cannot interfere with the field directly diffracted along the shadow boundary on the grazing ray.

The monotonic character of the correction to the geometrical optics cross section can be understood in terms of the results of Rice and Artmann. They showed that the shadow boundary is displaced toward or away from the shadow according as $u = 0$ or $\partial u/\partial \nu = 0$ on the cylinder. The magnitude of the displacement is proportional to $b^{1/3}k^{-2/3}$. Since each shadow boundary is shifted, the above type of result for the cross section is to be expected.

Finally, it should be pointed out that the methods used in this paper apply equally well to diffraction by three-dimensional objects and by objects with corners or edges. Diffraction by slits and by apertures of arbitrary shape has already been treated by these methods, and the results will be published soon.

APPENDIX I.

Determination of α

In order to determine α, we will first specialize (10) to the case of a circular cylinder of radius b. Let r and θ denote the polar co-ordinates of a point at which the field is to be evaluated and let the origin be on the cylinder axis. Then $s = (r^2 - b^2)^{1/2}$. The tangents to the circle from the point θ touch it at $\theta_1 = \theta - \cos^{-1}(b/r)$ and $\theta_2 = \theta + \cos^{-1}(b/r)$. We will consider only the ray tangent at θ_1. Then since arclength on the cylinder is given by $b\theta$, we have $t = b\theta_1 + \text{constant}$. The additive constant in t depends upon P_1, the point of initiation of the diffracted ray. Inserting these results into (10), and assuming that $\alpha(\tau)$ is constant, we obtain

$$u_{\text{diff}} \sim \frac{C}{(r^2 - b^2)^{1/4}}$$

$$\exp\left[ik(r^2 - b^2)^{1/2} + (ik - \alpha)b\left[\theta - \cos^{-1}\left(\frac{b}{r}\right)\right]\right]. \quad (45)$$

All factors which do not depend upon r or θ have been combined into the constant C.

For comparison, let us consider the solution

$$C'e^{i\nu\theta}H\nu^{(1)}(kr). \quad (46)$$

For any constants C' and ν, (46) is a solution of the reduced wave equation. We will show that it represents a wave traveling around a circular cylinder of radius b if ν is determined by

$$H\nu^{(1)}(kb) = 0. \quad (47)$$

The condition (47) is just the condition that the solution vanish on the cylinder. The function $H\nu^{(1)}$ is a Hankel function of the first kind of order ν.

In order to solve (47) we note that, since we will be interested in the asymptotic expansion of the solution for large k, the argument of the Hankel function will be large. Therefore we will replace the Hankel function by the first term in its asymptotic expansion for large argument. It will turn out that ν will also be large, so we will select an expansion valid when both the order and argument are large of the same order, i.e., when

$$|(\nu - kr)/kr| << 1.$$

Such an expansion is the so-called tangent approxima-
tion.[11]

$$H_v^{(1)}(kr) \sim \frac{2^{5/4} e^{-i\pi/4}}{\pi^{1/2}(kr)^{1/4}(v - kr)^{1/4}}$$

$$\cos\left\{\frac{\pi}{4} - \frac{i}{3}\left[2(v - kr)(kr)^{-1/3}\right]^{3/2}\right\}. \quad (48)$$

If we use (48) in (47) we obtain the following values for v

$$v = kb + (kb)^{1/3}\tau_n. \quad (49)$$

In (49), τ_n denotes a zero of $\cos\{\pi/4 - i/3[2\tau]^{3/2}\}$. The
τ_n are given by

$$\tau_n = \tfrac{1}{2}[3\pi(n + \tfrac{3}{4})]^{2/3}e^{i\pi/3}, \qquad n = 0, 1, 2, \cdots. \quad (50)$$

A table of values of τ_n is given in [12] pp. 44–45, where
more accurate values based upon the Hankel approxi-
mation instead of (48), are also given.

We have found that (46) is a solution of the reduced
wave equation which vanishes on the cylinder if v is
given by any one of the values (49). The asymptotic
form of this solution near the cylinder is obtained by
using (48) and (49) in (46) and noting that, except
where $r - b$ is involved, r may be replaced by b. We thus
obtain for r near b,

$$C'e^{iv\theta}H_v^{(1)}(kr)$$

$$\sim \frac{C'2^{5/4}\exp\left[-i\pi/4 + ikb\theta + i(kb)^{1/3}\tau_n\theta\right]}{\pi^{1/2}(kb)^{1/4}[kb + (kb)^{1/3}\tau_n - kr]^{1/4}}$$

$$\cos\left\{\frac{\pi}{4} - \frac{i}{3}\left[2(kb + (kb)^{1/3}\tau_n - kr)(kb)^{1/3}\right]^{3/2}\right\}. \quad (51)$$

At points r not near the cylinder we must use, instead
of (48), an expansion appropriate to Hankel functions
of large order and large argument, with the argument
larger than, and not nearly equal to, the order. Such an
expansion.[13] When inserted into (46) it yields for $r > b$,

$$C'e^{iv\theta}H_v^{(1)}(kr) \sim C'\sqrt{\frac{2}{i\pi k}}$$

$$(r^2 - b^2)^{-1/4}\exp\left\{ik(r^2 - b^2)^{1/2}\right.$$

$$\left. + i[kb + (kb)^{1/3}\tau_n]\left[\theta - \cos^{-1}\left(\frac{b}{r}\right)\right]\right\}. \quad (52)$$

Upon comparing (52) with (45) we see that the two
expressions agree exactly if the constant C' is chosen
appropriately and if

$$\alpha = -ik^{1/3}b^{-2/3}\tau_n. \quad (53)$$

---

[13] A. Erdelyi, "Higher Transcendental Functions," McGraw-Hill
Book Co., Inc., New York, N. Y., vol. II; 1953.

APPENDIX II
Determination of $B(P_1)$

In accordance with the method described above, we
will determine $B(P_1)$ by first specializing (11a) to apply
to a circular cylinder. We first note that in this case
$\rho_1 = \rho_2 = (R^2 - b^2)^{1/2}$ where R is the distance of the
source from the center of the cylinder. Furthermore, if
the exact incident wave is $(i/4)H_o^{(1)}(k\rho)$ then

$$A_o = \sqrt{\frac{i}{8\pi k}}. \quad (54)$$

We now use these values of ρ_1, ρ_2 and A_o in (11),
and assume the source to be at $\theta = 0$. Then we obtain
for the field diffracted by a circular cylinder upon which
a cylindrical wave is incident, the result

$$u_{\text{diff}} \sim \sqrt{\frac{i}{8\pi k}} \frac{e^{ik[(R^2-b^2)^{1/2}+(r^2-b^2)^{1/2}]}}{(R^2 - b^2)^{1/4}(r^2 - b^2)^{1/4}} \sum_{n=o}^{\infty} B^2{}_n(b)$$

$$\left\{e^{i\left[kb+(kb)^{1/3}\tau_n\right]}\theta - e^{i\left[kb+(kb)^{1/3}\tau_n\right](2\pi-\theta)}\right\}$$

$$\frac{e^{-i[kb+(kb)^{1/3}\tau_n][\cos^{-1}(b/r)+\cos^{-1}(b/R)]}}{1 - e^{2\pi i\left[kb+(kb)^{1/3}\tau_n\right]}}. \quad (55)$$

The result (55) agrees exactly with that of Franz[7]
based upon asymptotic expansion of the exact solution
of this problem, provided that $B_n(b)$ is given by (14).

APPENDIX III
Field Near the Cylinder

The diffracted field (10) which represents one of the
modes in (19), reduces to (45) for a circular cylinder.
Near the cylinder r is nearly equal to b and (45) becomes
in terms of s and θ

$$u_{\text{diff}} \sim \frac{C}{\sqrt{s}}\exp\left\{iks + i\left[kb + \tau_n(kb)^{1/3}\right]\left[\theta - \frac{s}{b}\right]\right\}. \quad (56)$$

The exact field (46) is given by (47) near the cylinder.
In terms of s and θ (47) becomes

$$C'e^{iv\theta}H_v^{(1)}(kr) \sim C'\sqrt{\frac{2}{\sqrt{i\pi k}}}\sqrt[4]{\frac{8}{b}}\frac{e^{ikb\theta+i(kb)^{1/3}\tau_n\theta}}{(b^{1/3}k^{-2/3}\tau_n - s^2/2\,b)^{1/4}}$$

$$\cos\left\{\frac{\pi}{4} - \frac{i}{3}(kb)^{1/2}\left[2(kb)^{1/3}\tau_n - \frac{ks^2}{b}\right]^{3/2}\right\}. \quad (57)$$

We note that although (56) becomes infinite as s
tends to zero, (57) does not become infinite. Therefore
near the cylinder (57) should be used instead of (56).
The ratio of (57) and (56) is just the factor by which (56)
must be multiplied to make it coincide with (57), and
thus remain finite. We now assume that this same factor,
which has been deduced for the case of a circular cylinder,
applies to any cylinder. Applying this factor to each
term of (19) then yields (27). In the same way (20) can
be modified to apply near the cylinder.

Near-Field Corrections to Line-of-Sight Propagation

A. D. WHEELON†

Summary—This study considers the line-of-sight propagation of electromagnetic waves in a turbulent medium. Interest here centers on the received signal's phase stability. The field equation describing propagation through a region characterized by random dielectric fluctuations is first developed. Solutions of this equation which represent the scattered field are derived with ordinary perturbation theory. These solutions are next used to calculate the rms phase error for an arbitrary path in the troposphere. This approach includes both a three-dimensional and near-field description for the multipath, scattered amplitudes, thereby overcoming the limitations of previous treatments. The phase correlation between signals received on two parallel transmission paths is derived last to illustrate the role of overlapping antenna beams.

I. Introduction

SCIENTISTS and engineers alike have displayed considerable interest in the recent success of microwave communication links operating well beyond the horizon. The propagation mechanism is presumably a scattering from turbulent fluctuations of the troposphere's dielectric constant.[1,2] As do all natural phenomenon, however, this scattering presents both opportunities and limitations. I shall focus this study on the restrictions which such fluctuations impose on ordinary line-of-sight propagation.

The reality of trophosphere scatter forces one to abandon the usual idealization of line-of-sight propagation to vacuum transmission. One must now anticipate that both phase and angle-of-arrival scintillations will arise from random variations of the atmosphere's dielectric constant. Such effects are evidently of prime importance when one asks for the ultimate accuracy of radio location schemes and high speed communication systems. The phase fluctuations on a single tropospheric path were first estimated using a rectilinear transmission (*i.e.*, geometrical optics) approximation[3] as,

$$\langle \alpha^2 \rangle \simeq \pi^2 \langle \Delta\epsilon^2 \rangle \frac{Ll_o}{\lambda^2} . \tag{1}$$

Where $\langle \Delta\epsilon^2 \rangle$ is the mean-square excursion of the dielectric constant from unity, l_o is the turbulence scale length, L denotes the transmission distance and λ is the incident radiation's wavelength. The contributions of multipath signals scattered by off line-of-sight turbulence structures can only be calculated with a full three-dimensional description of the scattered field. This extension was made in a second study[4] using the cross

section approximation. This latter approach relies on asymptotic expansions for the outgoing waves and is valid only when one measures the field at great distances from the scattering centers—as in scatter communication.

Line-of-sight propagation, however, is influenced by all turbulent "blobs" on the path, including those just next to the receiving antenna. The following development removes the distant field restriction by providing a field description valid throughout the transmission volume. Calculations of phase error and phase correlation are based on an expansion of the total field in powers of the dielectric fluctuations; corresponding to treating single, double, triple, etc. scattering as successive approximations. The incident field is assumed to be a simple plane wave and the received signal is to be measured with a moderately well-defined antenna pattern. Using principally the single scattering approximation, analytic expressions are then derived for the signal stability on an arbitrary path.

We shall frequently wish to facilitate and illustrate our calculations with typical turbulence parameters. Direct measurements of the troposphere's dielectric excursions with an airborne microwave refractometer[5] give several parts in 10^6. The associated scale length seems to vary between one and five hundred feet. For continuing reference, we shall choose $l_o = 200$ feet and $\langle \Delta\epsilon^2 \rangle = 10^{-12}$, quite arbitrarily. Herbstreit and Thompson will subsequently describe a series of most interesting experiments which were established to measure line-of-sight signal variations directly.[6] The path length in these experiments is almost 20,000 feet and the principal transmitter frequency 1,000 mc. We shall use these values in concert with the above to provide theoretical estimates for the rms phase errors on a single path and the phase correlation between parallel paths.

II. The Propagation Equation

One may represent the troposphere's electromagnetic properties by a single phenomenological parameter –its dielectric constant. We shall therefore concentrate our first effort on developing the field equations which describe propagation through a dielectric medium. Let us decompose the time and space-varying dielectric constant into its mean (unit Gaussian) value and a stochastic function to describe turbulent fluctuations of the medium:

† Ramo-Wooldridge Corp., Los Angeles 45, Calif.
[1] H. G. Booker and W. E. Gordon, "A theory of radio scattering in the troposphere," Proc. IRE, vol. 41, p. 401; April, 1950.
[2] F. Villars and V. F. Weisskopf; *Phys. Rev.*, vol. 94, p. 232; 1954.
[3] R. B. Muchmore and A. D. Wheelon, "Line-of-sight propagation phenomenon, I. direct ray." Proc. IRE, vol. 43, pp. 1437–1449; October, 1955.
[4] R. B. Muchmore and A. D. Wheelon, "Line-of-sight propagation phenomenon, II. scattered component." Proc. IRE, vol. 43, pp. 1450–1459; October, 1955.

[5] C. M. Crain, A. W. Straiton, and C. E. von Rosenberg, "A statistical survey of atmospheric index-of-refraction variation," Trans. IRE, vol. AP-1, no. 2, p. 43; October, 1953.
[6] J. W. Herbstreit and M. C. Thompson, "Measurements of the phase of signals received over transmission paths with their electrical lengths varying as a result of atmospheric turbulence," in this issue. Also Proc. IRE, vol. 43, pp. 1391–1401; October, 1955.

$$\epsilon(r, t) = 1 + \Delta\epsilon(r, t) . \tag{2}$$

The anticipated deviations from plane wave, phase stable transmission arise from the random variation $\Delta\epsilon \simeq 10^{-6}$

The starting point for any realistic description of electromagnetic propagation is Maxwell's equations. We shall consider these relations in a region characterized by unit permeability and the dielectric constant (2). In the lower atmosphere permanent charge and current densities are both sensibly absent, so that

$$\nabla x H = \frac{1}{c} \frac{\partial D}{\partial t} \tag{3}$$

$$\nabla x E = -\frac{1}{c} \frac{\partial B}{\partial t} \tag{4}$$

$$\nabla \cdot B = 0 \qquad \nabla \cdot D = 0 , \tag{5}$$

where,

$$B = H \quad \text{and} \quad D = \epsilon E . \tag{6}$$

The magnetic field may be eliminated in this system to yield:

$$\nabla^2 E - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \; (\epsilon E) = \nabla(\nabla \cdot E) . \tag{7}$$

Time derivatives here operate principally on E, since the atmospheric fluctuations represented by $\Delta\epsilon$ are considered very active if they rearrange themselves in a fraction of a second. If we denote the propagation constant by $K = \omega/c = 2\pi/\lambda$, the resulting equation is

$$[\nabla^2 + K^2(1 + \Delta\epsilon)]E = \nabla(\nabla \cdot E) . \tag{8}$$

We recognize the anomolous refraction effect in $K^2\Delta\epsilon$. The phase scintillations produced in a plane wave by this term have been investigated previously.[3] This same term's ability to scatter energy completely out of the primary (plane) wave implies that volume elements which do not lie directly on the rectilinear path can also contribute to the received signal's stability. Our interest will center on this latter problem. Before we can proceed, however, we must provide a brief interpretation for the divergence term $\nabla(\nabla \cdot E)$ in (8). If we combine (6) and (7), one finds,

$$\nabla \cdot E = -\frac{1}{\epsilon} E \cdot \nabla\epsilon \tag{9}$$

so that the gradient of this expression becomes:

$$\nabla(\nabla \cdot E) = \frac{\nabla\epsilon}{\epsilon} \left(E \cdot \frac{\nabla\epsilon}{\epsilon} \right) - \frac{1}{\epsilon} (E \cdot \nabla)\nabla\epsilon$$
$$- \frac{1}{\epsilon} (\nabla\epsilon \cdot \nabla)E - \frac{1}{\epsilon} \nabla\epsilon x(\nabla x E) .$$

Since each of these terms is already of order $\Delta\epsilon$, one may replace E with the incident plane wave $E_o = E_o j e^{iKz}$.

$$\nabla(\nabla \cdot E) \simeq E_o e^{iKz} \cdot \left\{ \bar{k} \frac{1}{\epsilon} \frac{\partial\epsilon}{\partial y} iK + \frac{1}{\epsilon^2} \frac{\partial\epsilon}{\partial y} \nabla\epsilon + \frac{1}{\epsilon} \frac{\partial}{\partial z} \nabla\epsilon \right\} .$$

Squares and derivatives of $\nabla\epsilon$ may usually be neglected in comparison with the normal gradient.

$$\nabla(\nabla \cdot E) \simeq \bar{k} E_o e^{iKz} \cdot iK \frac{1}{\epsilon} \frac{\partial\epsilon}{\partial y} . \tag{10}$$

This vector is directed along the incident field's propagation path, so that its composition with the primary wave distorts the phase front. This is simply equivalent to bending the incident beam, and its dependence on the normal gradient of ϵ provides contact with a similar result of geometrical optics. We must now contrast (10) with the anomolous refractive term in (8). The ratio of their magnitudes is $R = 2\pi\lambda/\Delta\epsilon \cdot \partial\epsilon/\partial y$, so that one may disregard the divergence term in (8) unless $\Delta\epsilon$ changes by its rms value in one wave length. Correlations of the major turbulence-sponsored dielectric excursions over hundreds of meters[7] ensure that this is quite a good assumption for radio frequencies. The central result of our discussion is the familiar linear propagation equation

$$[\nabla^2 + K^2(1 + \Delta\epsilon)]E = 0. \tag{11}$$

III. Solution for the Scattered Field

We shall now turn our attention to solving the propagation equation (11) for the scattered field. This task is complicated by the appearance of the stochastic variable $\Delta\epsilon$. We shall proceed as though this were an ordinary function of position, taking ensemble averages only at the end to recover the solution's statistical properties. We shall assume that the $\Delta\epsilon$ represent a Gaussian process with zero mean. The ensemble average of the product of n such variables may thus be expressed in terms of the space correlation function and mean-square dielectric fluctuation alone.

$$\langle\Delta\epsilon\rangle = 0$$
$$\langle\Delta\epsilon(r_1)\Delta\epsilon(r_2)\rangle = \langle\Delta\epsilon^2\rangle C(r_1 - r_2)$$
$$\langle\Delta\epsilon(r_1)\Delta\epsilon(r_2)\Delta\epsilon(r_3)\rangle = 0$$
$$\langle\Delta\epsilon(r_1)\Delta\epsilon(r_2)\Delta\epsilon(r_3)\Delta\epsilon(r_4)\rangle$$
$$= \langle\Delta\epsilon^2\rangle^2 \cdot [C(r_1 - r_2)C(r_3 - r_4)$$
$$+ C(r_1 - r_3)C(r_2 - r_4) + C(r_1 - r_4)C(r_2 - r_3)] \tag{12}$$
$$\dots \text{etc.}$$

We shall frequently wish to illustrate our results with concrete examples. To provide a consistent basis for comparison, we chose the Gaussian model:

$$C(R) = e^{-R^2/l_0^2} \tag{13}$$

to represent the atmosphere's turbulent aspects. The "blob coupling constant" $\langle\Delta\epsilon^2\rangle$ is a number like 10^{-12}, and provides a convenient expansion parameter.

Iterative Solution

To solve our fundamental field equation (11), we shall first adopt the philosophy of the quantum mechanical Born approximation. Its basic idea is to remove the term $k^2\Delta\epsilon$ to the right-hand side and use plane wave ($\Delta\epsilon = 0$) solutions to initiate an iterative solution. The result is a power series expansion in $\Delta\epsilon$; the zero order

[7] Detailed descriptions of turbulent structures indicate that an entire spectrum of blob sizes is present, each with its characteristic dielectric excursion ($\Delta\epsilon \simeq l^{2/3}$). This corresponds to an eddy decay scheme in which the initiating turbulent blobs subdivide until frictional forces just absorb their energy (see ref. 2). Although this lowest state has a scale size of the order of millimeters and is thus comparable with a wavelength, its intensity is sufficiently reduced to make it a small effect.

term of which is simply the incident wave. The term linear in $\Delta\epsilon$ describes single scattering; the quadratic, double scattering, and so on. We first recognize that the solution of (11) characterizes the *total* field E_T.

$$[\nabla^2 + K^2]E_T = -K^2\Delta\epsilon E_T. \qquad (14)$$

This equation may be inverted directly by introducing Green's function and the homogenous solution added to give:

$$E_T(R) = E_o(R) + K^2\int_V d^3r G(R, r)\Delta\epsilon(r)E_T(r), \qquad (15)$$

where the integration region V includes all volume elements which contribute to the scatter. We separate the total electric field into the incident plane wave E_0 and a scattered component E_s. The integral term in (15) evidently describes this scattered field. In the first approximation, E_T in the integrand is adequately given by the incident field, so that if ζ denotes the angle between E_o and the scattering direction, the amplitude of E_s is simply:

$$E_s(R) = K^2\int_V d^3r G(R, r)\Delta\epsilon(r)E_o(r) \, \text{Sin} \, \zeta + 0(\Delta\epsilon^2). \qquad (16)$$

The Green's function appropriate to our problem satisfies,

$$[\nabla^2 + K^2]G(R, r) = -\delta(R - r) \qquad (17)$$

and the outgoing wave condition of Sommerfeld, $\lim_{r\to\infty}[\partial/\partial r - iK]G(R, r) = 0$. Explicit and transform representations for $G(R, r)$ are:

$$G(R, r) = \frac{1}{4\pi}\frac{e^{iK|R-r|}}{|R - r|} \qquad (18)$$

$$G(R, r) = \frac{-1}{(2\pi)^3}\int d^3p \cdot \frac{e^{ip.(R-r)}}{K^2 - p^2 + i\epsilon} \qquad (19)$$

respectively. The second form incorporates a small imaginary term which directs one round the denominator's poles so as to represent outgoing waves.

1. *The Cross Section Approximation.* When one measures the field only at very great distance from the scattering volume V, one may employ the cross section approximation. This approach exploits an asymptotic expansion for the Green's function in (16) and is particularly well-suited to "scatter propagation" calculations, in which the common beam transmission volume V is midway between the sender and receiver.[1] If we denote the incident and scattered propagation vectors by K_o and K_1 respectively, it follows that

$$G(R, r) = \frac{e^{iKR}}{R} \cdot e^{-iK_1.r} + 0(1/R^2). \qquad (20)$$

If the incident plane wave is written $E_o = E_o e^{iK_0 r}$ one may compute the total power scattered to R with the aid of (16) as,

$$|E_s(R)|^2$$
$$= \frac{|E_o|^2}{R^2} \cdot \frac{K^4}{16\pi^2}\int_V d^3r_1\int_V d^3r_2 e^{iq.(r_1-r_2)}\Delta\epsilon(r_1)\Delta\epsilon(r_2)$$
$$\sin(\zeta_1)\sin(\zeta_2), \qquad (21)$$

where $q = K_1 - K_o$ and $|K_1| = |K_o| = K = 2\pi/\lambda$. This solution depends on the random $\Delta\epsilon$, so that one must yet average both sides on their ensemble. If one uses the second of relations (12) and a simple change of variable, the customary expression for the scattering cross section per unit volume, per unit solid angle emerges.

$$\sigma = \frac{R^2}{V}\left\langle\left|\frac{E_s(R)}{E_o}\right|^2\right\rangle$$
$$= \frac{K^4}{16\pi^2}\langle\Delta\epsilon^2\rangle\int d^3\rho e^{iq.\rho} C(\rho) \sin^2(\zeta) \qquad (22)$$

The cross section corresponding to the Gaussian model (13) is

$$\sigma = \frac{\sqrt{\pi}}{8\lambda}\langle\Delta\epsilon^2\rangle \sin^2(\zeta)\left(\frac{2\pi l_o}{\lambda}\right)^3$$
$$\exp -\left(\frac{2\pi l_o}{\lambda}\right)^2 \sin^2(\theta/2), \qquad (23)$$

where θ is the angle included between K_o and K_1. Since $l_o >> \lambda$ at radio frequencies, the scattered power is concentrated very strongly in the forward direction; the cone of "significant energy" having a half opening of $\lambda/2\pi l_o$ radians.

The scattering cross section finds several important applications. For instance, one may now compute the ratio of scattered to direct powers for line-of-sight propagation as

$$\frac{P_s(R)}{P_o} \simeq \int' dv \frac{\sigma}{r^2}, \qquad (24)$$

where the primed volume integration restricts that process to the receiving antenna's acceptance pattern. A conical antenna with half angle $\beta << 1$, looking through L meters of uniform turbulence described by (23) measures

$$\frac{P_s(R)}{P_o} \simeq \sqrt{\pi} \cdot \frac{Ll_o}{\lambda^2}\pi^2\langle\Delta\epsilon^2\rangle$$
$$\sin^2(\zeta)\left[1 - e^{-\left(\frac{\pi l_o\beta}{\lambda}\right)^2}\right]. \qquad (24)$$

Such estimations are always marred by the restriction that one measures this power only at great distances from all scattering centers. It may sound a bit paradoxical that we treat the near field of the blobs so carefully, while using a mere cutoff for the antenna pattern. Since the blobs are probably ten or more times as large as the receiving dishes, however, their far-field angular distribution is not realized so soon and we are evaluating the largest effect first. Detailed theory of receiver-medium coupling will be published soon.

2. *Symmetrical Theory.* We shall now turn to a fuller description of the scattered field. The power received at R may be computed without resorting to an asymptotic expansion for $G(R, r)$. In fact, one has only to square (16) and average directly over the $\Delta\epsilon$.

$$\left\langle \frac{P_s(R)}{P_o} \right\rangle = K^4 \langle \Delta \epsilon^2 \rangle \int_V d^3 r_1 \int_V d^3 r_2 G^*(R, r_1) G(R, r_2)$$

$$\cdot e^{iK_o \cdot \overline{(r_2 - r_1)}} C(|\overline{r_1 - r_2}|) \sin(\zeta_1) \sin(\zeta_2). \quad (25)$$

The correlation function now knits the two volume integrations together on an equal footing and thereby retains the manifest symmetry of the problem. The cross section approach averages the power scattered from a region of order l_o^3 about a particular point and then sums the contributions of all such points in the beam. Our new casting provides a satisfactory description of the singly-scattered field at all points in space. While its formal and conceptual advantages are evident, the computational hazards have increased correspondingly. We shall postpone an exemplary calculation until we discuss phase fluctuations of the received signal—a quantity of considerably more interest.

Multiple scattering corrections to the field may be derived by pursuing the interaction procedure outlined above. This process is now complicated, since volume elements which do not lie in the receiver's beam may take part in the first $n-1$ scatterings. Strictly speaking, one should run the iterated volume integrals over all space for which $\langle \Delta \epsilon^2 \rangle$ exists. The vigorous concentration of scattered energy in the forward direction described by our cross section (3.12) assures us that such scattering combinations are not likely to be important. The doubly-scattered wave correction to (3.14) may be written out with the help of (3.1).

$$\left\langle \frac{P_s(R)}{P_o} \right\rangle$$

$$= K^4 \langle \Delta \epsilon^2 \rangle \int d^3 r_1 \int d^3 r_2 G^*(R, r_1) G(R, r_2) e^{iK_o \cdot (r_2 - r_1)} C(r_1 r_2)$$

$$+ [K^4 \langle \Delta \epsilon^2 \rangle]^2 \int_V d^3 r_1 \int_V d^3 r_2 \int d^3 r_3 \int d^3 r_4 G^*(R, r_1) G(R, r_2)$$

$$+ G^*(r_1 r_3) G(r_2 r_4) e^{i \overline{K_o (r_4 - r_3)}} \cdot [C(r_1 r_2) C(r_3 r_4)$$

$$+ C(r_1 r_3) C(r_2 r_4) + C(r_1 r_4) C(r_2 r_3)]. \quad (26)$$

The second-order term here is exceedingly small compared with the first; so that one is led to regard first-order perturbation theory as a sufficient tool. Actually the series is an asymptotic one, in the sense that it converges for the first L/l_o terms and then begins to diverge. The utility of such representations for actual calculations is well known[8] so that we shall not labor the point here.

IV. LINE-OF-SIGHT PHASE STABILITY

A quantity of central interest for line-of-sight transmission is the rms phase error. We shall imagine a plane wave to fall on a region of turbulent dielectric, as shown in Fig. 1. An elementary approach to a calculation of phase stability is to consider only the energy which is propagated along the antenna's axis.[3,9] The Laplacian

[8] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw Hill Book Co., Inc., New York, N. Y., vol. I, p. 434; 1953.

[9] P. G. Bergmann, , *Phys. Rev.*, vol. 70, p. 486; 1946.
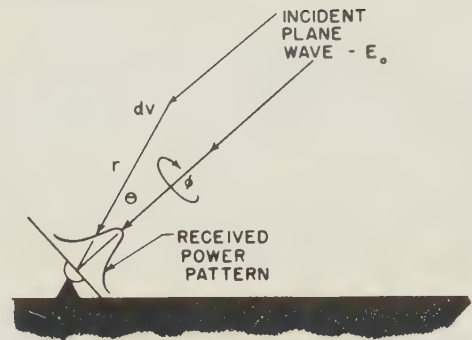
Fig. 1—Multipath propagation via single scattering.

operator in (11) may then be specialized to a single direction (s) and the WKB approximation[6] applied to give:

$$E(L) \simeq E(o) \frac{e^{iK \int_o^L ds\,(1 + \Delta \epsilon)^{1/2}}}{(1 + \Delta \epsilon)^{1/4}}. \quad (27)$$

The random phase fluctuations are recognized in this exponential form and correspond to alternate accelerations and retardations of the incident wave in traversing the stack of turbulent laminae. One now squares and averages this quantity over the $\Delta \epsilon$ with (12).

$$\langle \alpha^2 \rangle = \frac{\pi^2 \langle \Delta \epsilon^2 \rangle}{\lambda^2} \int_o^L ds_1 \int_o^L ds_2 C(|s_1 - s_2|). \quad (28)$$

With the Gaussian correlation (13), this becomes,

$$\langle \alpha^2 \rangle = \sqrt{\pi}\, \pi^2 \langle \Delta \epsilon^2 \rangle \frac{L \cdot l_o}{\lambda^2}, \quad (29)$$

although one should note that the result here is relatively insensitive to the actual correlation function employed.[3]
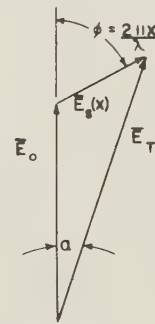
Fig. 2—Vector voltage diagram for superposition of multipath signals, x is the path difference.

The scattered field also contributes to the phase stability of the received signal. Off-axis turbulent blobs project a portion of their scattered energy into the receiving antenna. The greater propagation distances associated with these "kinked" paths implies phase delays for each such contribution. Their composition with the direct signal is illustrated by the vector voltage diagram of Fig. 2. The corresponding phase error is given by the ratio of the scattered field's out-of-phase component to the primary signal E_o.

$$\alpha \simeq \frac{1}{|E_o|} Im(E_s(R)). \tag{30}$$

One may now substitute the first-order approximation (16) for E_s and introduce the usual incident plane wave.

$$\alpha \simeq \frac{K^2}{4\pi} \int_V d^3r \frac{\Delta\epsilon(r)}{|R-r|} \sin[K\overline{(R-r)} + \overline{K_o}\cdot\overline{r}]. \tag{31}$$

To evaluate this expression, we choose the origin to coincide with the receiving antenna ($R = 0$) and erect a system of spherical co-ordinates on the antenna's axis, as in Fig. 1.

$$\alpha \simeq \frac{K^2}{4\pi} \int_o^L dr\, r \int_o^\beta d\theta \sin\theta \int_o^{2\pi} d\phi \Delta\epsilon(r)$$
$$\sin[Kr(1-\cos\theta)]. \tag{32}$$

Here L is the transmission distance and β denotes the half-angle of the reception beam.

The mean value of α vanishes identically, since $\langle\Delta\epsilon\rangle = 0$. Its mean square fluctuation survives and is given by a six-fold integral,

$$\langle\alpha^2\rangle = \frac{K^4}{16\pi^2} \langle\Delta\epsilon^2\rangle \int_o^L dr_1 r_1 \int_o^\beta d\theta_1$$
$$\sin\theta_1 \int_o^{2\pi} d\phi_1 \int_o^L dr_2 r_2 \int_o^\beta d\theta_2 \sin\theta_2 \int_o^{2\pi} d\phi_2$$
$$\sin[Kr_1(1-\cos\theta_1)] \sin[Kr_2(1-\cos\theta_2)]C(R_{12}). \tag{33}$$

R_{12} is the scalar distance between the two volume elements (1) and (2).

$$R_{12}{}^2 = r_1{}^2 + r_2{}^2$$
$$- 2r_1 r_2[\cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\phi_1 - \phi_2)]. \tag{34}$$

We shall assume a good directional receiver $\beta << 1$, so that small angle approximations may be used in (33) and (34). If one sets $r_1 = l_o x$ and $r_2 = l_o y$, with $M = L/l_o$ and $p = \pi l_o/\lambda$, it follows that

$$\langle\alpha^2\rangle = \frac{p^4}{\pi^2} \langle\Delta\epsilon^2\rangle$$

$$\int_o^M dx\, x \int_o^M dy\, y \int_o^\beta d\theta_1\theta_1 \int_o^\beta d\theta_2\theta_2 \int_o^{2\pi} d\phi_1 \int_o^{2\pi} d\phi_2$$
$$\sin(px\theta_1{}^2) \sin(py\theta_2{}^2)e^{-(x-y)^2}e^{-xy}[\theta_1{}^2+\theta_2{}^2$$
$$- 2\theta_2\theta_1\cos(\phi_1-\phi_2)]. \tag{35}$$

This integral is a formidable one and a Monte Carlo machine sampling procedure immediately suggests itself. Automatic digital computation is made extremely difficult, however, by the very rapid oscillations of the phase (sine) terms and encourage one to attempt the integrations by analytical devices.

We shall evaluate the multiple integral (35) in several interesting limits. The pertinent quantities are βM and βp. The first describes the ratio of the beam's largest (conical) radius to the scale size, and is essentially a measure of the number of blobs in the receiving beam. We found in (23) that $1/p = \lambda/\pi l_o \simeq \gamma$ represents the cone angle of significant scattered energy from a single blob. Those blobs which project a measurable amount of energy into the receiver must therefore lie within a beam of half angle γ. The effective receiver cone is then the lesser of these angles, so that $p\beta$ must be considered carefully.

1. $\beta M < 1$

When this condition obtains one may replace the entire second exponential in (35) by unity and perform the angular integrations analytically.

$$\langle\alpha^2\rangle = 4p^2\langle\Delta\epsilon^2\rangle \int_o^M dx \int_o^M dy\, e^{-(x-y)^2}$$
$$\sin^2\left(\frac{px\beta^2}{2}\right) \sin^2\left(\frac{py\beta^2}{2}\right). \tag{36}$$

In the usual physical situation, the scattering beam is much smaller than the receiver's beam ($p\beta > 1$), so that p and not β should appear in the answer. The trigonometric terms in (36) each oscillate very rapidly ($p\beta > 1/M\beta$) with average values of $\frac{1}{2}$. The resulting integrations give the rectilinear result,

$$\langle\alpha^2\rangle = \sqrt{\pi}\, Mp^2\langle\Delta\epsilon^2\rangle$$
$$= \sqrt{\pi}\pi^2 \frac{Ll_o}{\lambda^2} \langle\Delta\epsilon^2\rangle. \tag{37}$$

To understand this coincidence physically, we note that the greatest conical radius seen by the receiver is less than one scale length, so that the beam sees only small sections of whole blobs. This sequence of turbulent "plugs" appears as a stack of almost uniform turbulent laminae to the receiver and the problem is substantially one dimensional. Bremmer's derivation[10] of the first WKB approximation from a continuous single transmission (forward scattering in one dimension) picture provides the necessary contact between the two methods of calculation.

The opposite limit, $p\beta = \beta/\gamma < 1$ can be evaluated directly by using small argument expansion for the sine terms. This corresponds physically to being always in the near field of *both* the blobs and antenna. Close to the receiver, however, one cannot use a simple (β) cutoff for the antenna pattern, so that we shall omit this case.

2. $\beta M > 1$

The more difficult problem occurs when the receiver sees several blobs across the width of the beam. This is the situation in the Cheyenne Mountain experiments[6] for which $\beta M \simeq 10$. Scintillation then arises from a genuine three-dimensional scattering mechanism and the angular co-ordinates in the second exponential of (35) cannot be ignored. The azimuthal integrations may be performed to give $4\pi^2 I_o(2xy\theta_1\theta_2)$, which in turn we approximate by its asymptotic form. (The rather more accurate interpolation formula, $I_o(u) = e^u/\sqrt{1+2\pi u}$ is not required here since the factor $xy\theta_1\theta_2$ handles the singularity at the origin.)

[10] H. Bremmer, paper in "Theory of Electromagnetic Waves," Interscience Publishers, New York, N. Y., p. 105; 1951.

$$\langle\alpha^2\rangle \simeq \frac{2p^4}{\sqrt{\pi}}\langle\Delta\epsilon^2\rangle \int_0^M dx \int_0^M dy \int_0^\beta d\theta_1 \int_0^\beta d\theta_2$$

$$\sqrt{xy\theta_1\theta_2}\,e^{-(x-y)^2}\sin[px\theta_1^2]\sin[py\theta_2^2]e^{-xy(\theta_1-\theta_2)^2}. \quad (38)$$

The principle contribution to the x, y integrals comes from the diagonal $x = y$. We introduce a factor $\sqrt{\pi}$ to account for the effective cross span of this ridge,[11]

$$\langle\alpha^2\rangle \simeq p^4(\Delta\epsilon^2) \int_0^M dx\,x \int_0^\beta d\theta_1 \int_0^\beta d\theta_2 \sqrt{\theta_1\theta_2}\,e^{-x^2(\theta_1-\theta_2)^2}$$

$$\{\cos[px(\theta_1-\theta_2)(\theta_1+\theta_2)] - \cos[px(\theta_1^2+\theta_2^2)]\}. \quad (39)$$

The experimentally observed increase of phase scintillation with distance leads one to expect (α^2) to vary about as M. The x integral must thus depend on running over the large range (M), and not on gathering contribution from small x. When x is large, however, the exponential term is very small unless θ_1 and θ_2 are nearly equal. We may estimate this coalescence directly by rotating the co-ordinates $\pi/4$ radians: $u = \theta_1 - \theta_2$ and $v = \theta_1 + \theta_2$.

$$\langle\alpha^2\rangle \simeq \frac{p^4}{2}\langle\Delta\epsilon^2\rangle \int_0^M dx\,x \int_0^\beta du \int_u^{2\beta-u} dv\,\sqrt{v^2-u^2}\,e^{-x^2u^2}$$

$$\cos(pxuv) - \cos\left[px\frac{u^2+v^2}{2}\right]. \quad (40)$$

For large x, the principle contribution to the u integral comes from small values, so that one may eliminate u from the v limits, square root, and second cosine. With the condition $\beta M > 1$ one may now relax β to infinity in the u integration to find,

$$\langle\alpha^2\rangle \simeq \frac{p^4}{2}\langle\Delta\epsilon^2\rangle \int_0^M dx\,x \int_0^{2\beta} dv\,v$$

$$\int_0^\infty du\,e^{-x^2u^2}\left[\cos(pxuv) - \cos\left(\frac{pxv^2}{2}\right)\right]. \quad (40)$$

The second cosine oscillates rapidly and gives zero integral, if $p\beta^2 M \gg 1$. The remaining integrations give

$$\langle\alpha^2\rangle = \tfrac{1}{2}p^2 M\sqrt{\pi}(1-e^{-p^2\beta^2})\langle\Delta\epsilon^2\rangle, \quad (41)$$

which is exactly the result predicted by cross section theory.[4] In the limit $p\beta > 1$, it differs from the rectilinear result (30) only by a factor $\frac{1}{2}$.

When the scattering angle $\gamma = 1/p$ is greater than the beamwidth, one derives the β-dependent result,

$$(\alpha^2) = \frac{\sqrt{\pi}}{2}\cdot\pi^4\,\frac{Ll_o^3}{\lambda^4}\,\beta^4(\Delta\epsilon^2). \quad (42)$$

This stronger dependence on the frequency is of considerable importance for high resolution antennas or small turbulence scale lengths. Other inequalities amongst the transmission parameters may be evaluated from (35) by similar integration techniques.

The approach given here has the advantage of providing an accurate near-field description. Since blobs which stand close both to the axis and receiver have small path differences and large scattered amplitudes,

[11] Using the formula.

$$\int_0^M dx \int_0^M dy f(x,y)e^{-(y-x)^2} \simeq \sqrt{\pi}\int_0^M dx f(x,x) \quad \text{for} \quad M \gg 1.$$

their phase effect is undoubtedly quite large. Previous treatments[4] of this problem based on the cross section approximation suffer in this critical region. For line-of-sight propagation, the added computational burden of our symmetric formulation seems fully justified. The sufficiency of first-order perturbation theory is argued for by the discussion of (26).

V. PHASE CORRELATION

The use of two-spaced antennas to observe the same transmitter has proven a powerful tool in tropospheric investigations.[6] The phase correlation between two antennas, spaced D meters apart normal to the transmission path, depends only on the ratio D/l_o. If both the single path phase error and phase correlation between two adjacent paths are measured simultaneously, one can determine the two phenomenological scattering parameters l_o and $\langle\Delta\epsilon^2\rangle$ at once.

To provide a theoretical estimate of this correlation, one may go to several levels of sophistication. The simplest is that employed in reference (3) to discuss two parallel, rectilinear paths. From Fig. 3 and (28),

$$\langle\alpha_1\alpha_2\rangle = \frac{\pi^2}{\lambda^2}\langle\Delta\epsilon^2\rangle \int_0^L ds_1 \int_0^L ds_2\,C(\sqrt{(s_1-s_2)^2+D^2});$$

$$(43)$$

so that with the turbulence model of (13),

$$\rho_{12} = \frac{\langle\alpha_1\alpha_2\rangle}{\langle\alpha^2\rangle} = e^{-D^2/l_o^2}. \quad (44)$$
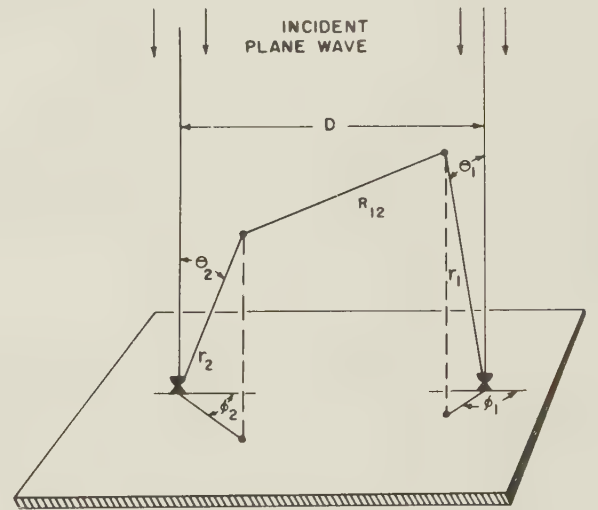
Fig. 3—Coordinates for calculation of phase correlation between parallel transmission paths.

The normalized phase correlation is thus a rapidly decreasing function of D/l_o. Such a treatment necessarily neglects the overlapping region of the two receiver's (diverging) antennas patterns; an effect which should tend to increase the correlation (44). A short investigation shows that the cross section approach precludes such a calculation by its very assumptions, so that we must immediately use the symmetrical theory.

One may properly include these common volume elements by simply averaging the product of two disjoint α's. If we establish a system of spherical co-ordinates on each receiver's pointing direction, one has from (32):

$$\langle \alpha_1 \alpha_2 \rangle = \frac{K^4}{16\pi^2} \langle \Delta\epsilon^2 \rangle \int_o^L dr_1 r_1 \int_o^\beta d\theta_1$$

$$\sin\theta_1 \int_o^{2\pi} d\phi_1 \int_o^L dr_2 r_2 \int_o^\beta d\theta_2 \sin\theta_2 \int_o^{2\pi} d\phi_2$$

$$\sin[Kr_1(1-\cos\theta_1)]\sin[Kr_2(1-\cos\theta_2)]e^{-R_{12}^2/l_o^2}. \quad (45)$$

The distance between typical volume elements in beam (1) and (2) is now a rather complicated function of all the variables.

$$R_{12}^2 = D^2 + r_1^2 + r_2^2$$

$$- 2r_1 r_2[\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_1 - \phi_2)]$$

$$- 2D[r_1\sin\theta_1\cos\phi_1 - r_2\sin\theta_2\cos\phi_2]. \quad (46)$$

The role of the overlapping conical antenna beams may be inferred from this form. The essential competition is between terms which are (at most) of order $(L\beta)^2$ and those of order $L\beta D$. If both are much less than l_o^2, the rectilinear result (44) is rightly regained. If $2L\beta$ is less than D, the two beams do not overlap and one expects few corrections to (44). It is only when both beams see common scattering volumes ($q = D/l_o < 2M\beta$) that a significant modification should arise. These conditions are represented mathematically by keeping the first and/or second square brackets in (46).

We again assume $\beta \ll 1$ and proceed with the evaluation of (46) according to the value of βM, utilizing the abbreviations introduced after (34). We shall actually calculate the phase correlation $\rho_{12} = \langle\alpha_1\alpha_2\rangle/\langle\alpha^2\rangle$ as a function of $q = D/l_o$, by dividing with the appropriate mean square value from section IV.

1. $\beta M < 1$

For small beamwidths, one may replace the factor of $r_1 r_2$ by unity in (46) and perform the azimuthal integrations directly. We focus our efforts on values of $p\beta > 1/\beta M$, so that with (37):

$$\rho_{12} = \frac{4p^2 e^{-q^2}}{\pi^2\sqrt{\pi}M} \int_o^M dxx \int_o^M dyy \int_o^\beta d\theta_1\theta_1 \int_o^\beta d\theta_2\theta_2 e^{-(x-y)^2}$$

$$\sin(px\theta_1^2)\sin(py\theta_2^2)I_o[2qx\theta_1]I_o[2qy\theta_2]. \quad (47)$$

Quite a large range of receiver spacings is covered by $q < 1/\beta M$, for which small argument expansion may be used for the Bessel functions and the integrals performed analytically. To first order in q^2,

$$\rho_{12} = \left[1 - q^2\frac{2M}{p}\sin(pM\beta^2)\right]e^{-q^2}. \quad (48)$$

Since $\beta M < 1$ and $p\beta >> 1$, M/p is less than unity and the calculated correction to (44) is small. This is in agreement with our previous arguments, since here $q > \beta M$. Larger values of q are quite rare, so that we proceed directly to the next limit.

2. $\beta M > 1$

When $\beta M > 1$, we may encompass a rather large range of q, while satisfying the condition for overlapping beams, $\beta M > q$. Both exponential terms in (46) must now be retained and the first azimuthal integration can be performed immediately.

$$\mu = \int_o^{2\pi} d\phi_1 \int_o^{2\pi} d\phi_2 e^{-2xy\theta_1\theta_2\cos(\phi_1-\phi_2)-2qx\theta_1\cos\phi_1+2qy\theta_2\cos\phi_2}$$

$$= 2\pi\int_o^{2\pi} d\phi_1 e^{-2qx\theta_1\cos\phi_1}$$

$$I_o[2y\theta_2\sqrt{x^2\theta_1^2 - 2qx\theta_1\cos\phi_1 + q^2}].$$

One may safely resort to the I_o function's asymptotic form and invoke the inequality $\beta M > q$ to expand the troublesome square root. The second integral can thus be done to give

$$\mu = \frac{2\pi^2}{\sqrt{\pi xy\theta_1\theta_2}} e^{2xy\theta_1\theta_2}I_o[2q(x\theta_1 + y\theta_2)], \quad (49)$$

which is independent of the order of integration employed. When this result is substituted into (45), the form is that for the mean-square phase shift, except for two new factors: e^{-q^2} and $I_o[2q(x\theta_1 + y\theta_2)]$. We proceed to collapse the $x - y$ integration as before and rotate the θ_i coordinates by $\pi/4$. With the approximations used in section IV and $p\beta > 1$,

$$\rho_{12} = \frac{p^2}{M\sqrt{\pi}} e^{-q^2}$$

$$\int_o^M dxx \int_o^{2\beta} dvvI_o[2qxv] \int_o^\beta due^{-x^2a^2}\cos(pxuv). \quad (50)$$

When x is large, the β limit of the u integration may be sent to infinity.

$$\rho_{12} \simeq \frac{1}{2}\frac{p^2}{M} e^{-q^2} \int_o^M dx \int_o^{2\beta} dvvI_o[2qxv]e^{-p^2v^2/4}.$$

The v integral may now be done analytically if one recalls that $p\beta > 1$, which allows 2β to be relaxed to infinity if $p > M$,

$$\rho_{12} \simeq \frac{e^{-q^2}}{M} \int_o^M dxe^{\frac{4q^2}{p^2}x^2}. \quad (51)$$

When q is less than one and $M < p$, one may expand the exponential to find,

$$\rho_{12} \simeq 1 - q^2\left(1 - \frac{4}{3}\frac{M^2}{p^2}\right) + 0(q^4). \quad (52)$$

Overlapping antennas patterns are thus seen to enhance the phase correlation over that which would be calculated

for parallel, disjoint paths. The result is independent of β, since the scattering cone again serves as the effective beamwidth.

When q is greater than one, the integration of (51) needs be performed numerically. Setting $x = Mz$ brings out the essential dependence on $2M/p$. Upper and lower bounds for the correlation are given by:

$$e^{-q^2} \leq \rho_{12} = e^{-q^2} \int_0^1 dz\, e^{z^2 q^2 (2M/p)^2} \leq e^{-q^2[1-(2M/p)^2]}, \quad (53)$$

and provide a very good estimate if $2M/p$ is small. When this ratio is one-sixth, the geometrical optics result (44) is verified to within ten per cent over most of the range on q. The integration scheme which leads to (51) breaks down when $2M/p$ reaches one and a new approach is required. The possible combinations of transmission parameters are too numerous to be dealt with exhaustively here. We content ourselves with having studied those cases which correspond to the propagation measurements mentioned at the outset. Further work should probably be directed toward correlation functions other than (13), so as to test the sensitivity of this broader theory to the assumed properties of the turbulence structure.

On the Scattering of Waves by an Infinite Grating*

VICTOR TWERSKY†

Summary—Using Green's function methods, we express the field of a grating of cylinders excited by a plane wave as certain sets of plane waves: a transmitted set, a reflected set, and essentially the sum of the two "inside" the grating. The transmitted set is given by $\psi_0 + 2\Sigma C_v G(\theta_v, \theta_0)\psi_v$, where the ψ's are the usual infinite number of plane wave (propagating and surface) modes; $G(\theta_v, \theta_0)$ is the "multiple scattered amplitude of a cylinder in the grating" for direction of incidence θ_0 and observation θ_v; and the C's are known constants. (For a propagating mode, C_v is proportional to the number of cylinders in the first Fresnel zone corresponding to the direction of mode v.) We show (for cylinders symmetrical to the plane of the grating) that

$$G(\theta, \theta_0) = g(\theta, \theta_0)$$
$$+ \ (\Sigma_v - \int dv)C_v[g(\theta, \theta_v)G(\theta_v, \theta_0)$$
$$+ \ g(\theta, \pi - \theta_v)G(\pi - \theta_v, \theta_0)],$$

where g is the scattering amplitude of an isolated cylinder. This inhomogeneous "sum-integral" equation for G is applied to the "Wood anomalies" of the analogous reflection grating; we derive a simple approximation indicating extrema in the intensity at wavelengths slightly longer than those having a grazing mode. These extrema suggest the use of gratings as microwave filters, polarizers, etc.

I. Introduction

The scattering of waves by gratings (i.e., structures which are periodic in one dimension) is of interest in many branches of physics and engineering. Gratings are used widely as spectrum analyzers and as "open reflectors," and such applications are more or less adequately treated by simple approximations of the boundary value problem arising from Maxwell's equations. Thus the presence of the spectra may be deduced from Fresnel theory,[1] and a "single scattering" treatment often suffices to predict the "modulation" introduced by a particular choice of grating elements.[2] Also, certain characteristics of the grating as a simple reflector follow from a "static approximation."[3]

However, aside from the fact that more rigorous expressions may be desirable for the above applications, there are certain phenomena which the simple approximations fail to predict. Such approximations indicate that continuous radiation resolved by a grating should yield an approximately uniform intensity-wavelength curve in a given spectral order. Thus they do not account for the "Wood anomalies"[4] (dark and bright,

narrow bands observed primarily for white light polarized perpendicular to the elements of certain reflection gratings), or for related phenomena which suggest the use of gratings as microwave filters, polarizers, etc.

Rayleigh[5] noted, in connection with Wood's original data, that "anomalies" might well occur for wavelengths having a spectral line in the plane of the grating (the "Rayleigh wavelengths"): at these wavelengths, the "secondary waves" are all in phase at an element (e.g., at normal incidence, the grating spacing equals an integral number of such wavelengths). Later measurements found that the intensity-wavelength curve of many anomalies had a sharp edge at a Rayleigh wavelength, followed by a minimum and maximum at longer wavelengths; other anomalies were more complex (as determined by the particular grating elements), and their appearance differed in the various spectral orders. These anomalies are "surface wave" effects, or, equivalently, "multiple scattering" effects; their analysis requires more careful treatment of the boundary value problem.

The field of an infinite grating excited by a monochromatic plane wave consists essentially of an infinite set of plane waves; some of these waves are "propagating modes," and these carry energy in specific directions; the remaining are "surface waves," which are exponentially damped normal to the plane of the grating. The existence of these waves follows from the periodicity of the structure; i.e., the field must be representable as a Fourier series, and this has been the starting point of most investigators; in particular, Rayleigh[6] first applied this procedure to the "corrugated grating," and Ignatowsky[7]

[1] F. M. Schwerd (1835); see E. Mach, "Physical Optics," Dover Press, New York, N. Y., p. 291; 1953.

[2] C. Schaefer and F. Reiche, *Ann. Phys.*, vol. 35, p. 817; 1911, gave a "single scattering" treatment of a grating consisting of a finite number of circular cylinders; i.e., they treated each cylinder as isolated from the remainder.

[3] H. Lamb, "Hydrodynamics," Dover Press, New York, N. Y., p. 537 ff; 1945, treated an acoustic wave and rigid circular cylinders. R. Gans, *Ann. Phys.*, vol. 61, p. 447; 1920, treated both polarizations for arbitrary conducting circular cylinders.

[4] R. W. Wood, *Phil. Mag.*, vol. 4, p. 396; 1902: vol. 23, p. 310; 1912, *Phys. Rev.*, vol. 48, p. 928; 1935. L. R. Ingersoll, *Astrophys. Jour.*, vol. 51, p. 129; 1920, *Phys. Rev.*, vol. 17, p. 493; 1921. J. Strong, *Phys. Rev.*, vol. 49, p. 291; 1936. C. H. Palmer, *Jour. Opt. Soc. Amer.*, vol. 42, p. 269; 1952. Pronounced anomalies for *E*-polarization perpendicular to the elements were investigated by all the above; apparently fainter "parallel anomalies," although mentioned by Wood, were investigated only by Palmer.

[5] Rayleigh, *Phil. Mag.*, vol. 14, p. 60; 1907.

[6] Rayleigh, *Proc. Roy. Soc.*, London, vol. A79, p. 399; 1907, and C. T. Tai, Tech. Rep. No. 28, Cruft Lab., Harvard University, Cambridge, Mass., 1948, treated perfectly conducting surfaces with "square wave" profiles, or, more generally, profiles which could be represented by Fourier series. This was extended to absorbing surfaces by W. Voigt *Göttinger Nachrichten*, pp. 40 ff, 1911, U. Fano, *Ann. Phys.*, vol. 32, p. 393; 1938, and K. Artmann, *Z. Phys.*, vol. 119, p. 529; 1942. More recently, B. A. Lipmann, Rep. 18-8, Nuclear Development Associates, White Plains, N. Y.; 1953, treated the perfect conductor by Green's function methods.

[7] W. v. Ignatowsky, *Ann. Phys.*, vol. 44, p. 369; 1914, considered the general transmission grating, showed that his result reduced to approximate forms previously obtained by others (e.g., Schaefer and Reiche, *op. cit.*), and then treated normal incidence for circular cylinders. The special case of "isotropic scatterers" (polarization parallel to fine perfect conductors) was treated in detail by W. Wessel, *Hochfreq. und Elektroak.*, vol. 54, p. 62; 1939, R. Honerjager, *Ann. Phys.*, vol. 6, p. 25; 1948, and W. Franz, *Z. Angew Phys.*, vol. 9, p. 416; 1949. A Green's function treatment of perfect conductors was given by J. Schmoys, *Jour. Opt. Soc. Amer.*, vol. 41, p. 324; 1951, and specialized to circular cylinders by W. Sollfrey and J. Schmoys, Rep. EM41, Math. Res. Group, New York University, New York; 1952. Green's function methods were also used by S. Karp, Progress Rep. No. 13, Math. Res. Group, New York University, New York; 1952. V. Twersky, *Jour. Appl. Phys.*, vol. 23, p. 1099; 1952, using essentially an iterated form of Ignatowsky's equations, treated the finite grating of circular cylinders heuristically. F. Reiche, Rep. EM-61, Inst. Math. Sci., Div. Electromag. Res., New York University, New York; 1953, showed that the representation for circular cylinders in terms of complete sets of cylindrical waves could be reduced to plane waves.

to a grating of arbitrary infinite cylinders. However, this is merely the start of the analysis; it is the amplitudes of these waves which must be determined. Rayleigh,[6] and others, obtained series representations of an originally infinite set of algebraic equations involving the mode amplitudes and the amplitudes of the Fourier components of the surface profile. Ignatowsky's[7] amplitudes were expressed in terms of the current distribution on one cylinder, and the equations for the current were approximated both analytically, and heuristically in series and in closed form. Schmoys,[7] Sollfrey,[7] and Lipmann[6] used the Schwinger variational approximation. Karp,[7] and Twersky[7] used heuristic physical approximations in order to express the mode amplitudes in closed form in terms of the scattering amplitude of an isolated cylinder.

In connection with the Wood anomalies, Rayleigh[6] treated a perfectly conducting grating, and found that his representation of the perpendicular polarized amplitudes diverged if there was a grazing mode; Fano[6] considered the same range for a nonperfectly conducting grating; Artmann[6] pointed out that Fano's result was unrealistic, and derived an alternative, convergent series representation. There is no doubt that Artmann's expressions for the maxima correspond to the maxima of the usual Wood anomalies; although Artmann does not consider the associated minima (lying between the Rayleigh wavelengths and the maxima), these may also be treated using Rayleigh's model. Fano also presented a general expression (suggested by a quantum mechanical analogy) which could be adjusted to fit the described anomalies, and he was also the first to stress the role of the surface waves. Twersky[7] gave an "orders of scattering" treatment of the anomalies for a finite grating (perfectly conducting semicylinders on a plane); the extrema were interpreted as occurring at wavelengths which optimally fulfilled the conditions that each order of scattering be a maximum, and that successive orders be either in or out of phase. Here a suggestion made by Wood,[4] and originally elaborated by Artmann,[6] was developed into a "vibration curve" method based on a discrete analog of the Fresnel integral.

In the present paper, although we do not attempt a comprehensive treatment of the anomalies, we derive a relatively simple expression indicating their primary features; here the interpretation is essentially in terms of surface waves. More generally, whereas previous writers expressed the mode amplitudes initially in terms of either an integral equation or an infinite set of algebraic equations, we derive a somewhat more tractable "mixed representation." Our procedure, which is essentially an extension of Karp's, is the following:

In Section II we review the two-dimensional Green's function treatment of an arbitrary cylinder. Here we define the "scattering amplitude of an isolated cylinder," $g(\theta, \theta_o)$, which specifies the far-field response in the direction θ to a wave incident in a direction θ_o; for later convenience we assume the cylinder has a symmetry plane parallel to its generator. This function g is assumed to be known, so that we may proceed directly

to the novel features of a grating composed of such cylinders (the grating having the same symmetry plane as its components).

In Section III we develop a variation of the usual Green's function treatment of the grating. (We impose the radiation condition on the wave scattered by an individual cylinder, rather than on the field scattered by the grating as a whole; it is shown that both procedures are equivalent.) We introduce the "multiple scattered amplitude of a cylinder in the grating," $G(\theta, \theta_o)$, and derive a plane wave representation of the total scattered field of the grating, such that the plane wave amplitudes equal $2C_v G(\theta_v, \theta_o)$. For the propagating modes, C_v is proportional to the number of elements in the first Fresnel zone (here two strips) of mode v; each mode is proportional to the wave scattered as a group by the elements in half the first Fresnel zone appropriate to that direction. The complete representation of the field consists of three sets of plane waves: a transmitted set, a reflected set, and essentially the sum of the two "inside the grating". (A form similar to this last set was first obtained explicitly by Lipmann.[6])

The aim of the paper is then to express G analytically in terms of g. For this purpose we represent the total field at the surface of a cylinder in the grating as a sum of three terms: (1) the incident plane wave; (2) the difference of a discrete set and its analogous continuous set of plane wave modes; and (3) the wave scattered by the cylinder. The continuous set of plane waves equals the scattered cylindrical wave; we subtracted the first and added the second in order to represent the total field at the scatterer as some collection of plane waves plus one scattered cylindrical wave. From this form (and the superposition principle) we may immediately represent $G(\theta, \theta_o)$ as the sum of $g(\theta, \theta_o)$ plus a set (over v) of terms of the form $g(\theta, \theta_v)C_v G(\theta_v, \theta_o)$. (An approximation of this result, with the entire set replaced by two grazing modes, was obtained originally by Karp.[7]) The set itself, the sum over propagating and surface modes *minus* the analogous integral, is what simplifies this inhomogeneous "sum-integral" equation for $G(\theta, \theta_o)$; i.e., the surface terms are in general negligible if the spacing is moderately large compared to wavelength; the exceptional cases, for which one or two surface modes are very large, correspond to the Wood anomalies.

Having derived a representation for G, we extend it (by the image method) to represent the corresponding scattering amplitude of a protuberance in a reflection grating. (Thus the reflection grating treated by Rayleigh[6] corresponds to reflecting the results for the transmission grating of perfectly conducting, rectangular cylinders.)

In Section IV, we derive certain scattering theorems fulfilled by the G's; these theorems and their single scattered anologs are used to motivate the analysis of Section V, and to check approximations. We apply the results to obtain approximations for the transmission and reflection coefficients of a grating for which only one or two modes propagate. However, the principal application is to the Wood anomalies of a perfectly conducting

reflection grating with symmetrical elements; the approximations obtained for polarization perpendicular to the elements indicate well defined extrema of the type mentioned previously; for parallel polarization, however, only a trend seems indicated. In general, the presence of an intense grazing mode seems to be the only explicit requirement for the appearance of the anomalies.

Thus our model is in essential accord with the data of Wood,[4] Ingersoll,[4] and Strong,[4] but does not account for the marked parallel effects observed by Palmer.[4] We believe that the discrepancy arises from our perfectly conducting ground plane; i.e., the field of an arbitrary protuberance and an incident wave polarized parallel to the ground plane must vanish in a direction parallel to the plane. Thus, since Palmer observed parallel anomalies only for gratings having a marked parallel polarized component at grazing, we infer that his gratings did not have a highly conducting surface; e.g., perhaps the bottoms of the grating grooves were uncoated, or poorly coated with metal. (Palmer's data should be directly interpretable by extending Artmann's[6] results for an imperfectly conducting grating to the case of parallel polarization.)

In Section VI we specialize the above general results to a grating of circular cylinders, and to the analogous reflection grating. We also give approximations of the transmission and reflection coefficients for perfect conductors with radii and spacing small compared to wavelength; our results for perpendicular polarization are different than those of Lamb and Gans,[3] our ratio of reflection coefficient to transmission coefficient for normal incidence being more than twice their value.

II. The Single Cylinder

The two-dimensional problem of the scattering of a plane wave (electromagnetic or acoustic) by a cylinder parallel to the z axis is specified by a solution of

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0, \qquad \nabla^2 = \partial_x{}^2 + \partial_y{}^2, \qquad (1)$$

subject to prescribed boundary conditions on the cylinder's surface. In the region external to the cylinder, the solution is to be of the form

$$\psi(\mathbf{r}) = \psi_o(\mathbf{r}) + u(\mathbf{r}), \qquad k = 2\pi/\lambda, \qquad (2)$$

where ψ_o is a plane wave

$$\psi_o(\mathbf{r}) = e^{ikx \cos \theta_0 + iky \sin \theta_0} = e^{ikr \cos (\theta - \theta_0)}, \qquad (3)$$

and where $u(\mathbf{r})$, the corresponding scattered wave, is to fulfill the radiation condition, i.e., $u(\mathbf{r})$ is to be an outgoing cylindrical wave as $r \to \infty$; the suppressed time factor is $e^{-i\omega t}$. The internal field, which we do not discuss explicitly, is to be bounded.

The wave function ψ is the velocity potential for the acoustic case of small amplitude vibrations; for the electromagnetic case, ψ equals either E_z or H_z. The boundary conditions on ψ follow (from the original acoustic and electromagnetic equations) after selection of the physical constants of the cylinder, and the direction of incident polarization. Once these are specified, we may construct auxiliary equations whose solution (at least in principle) lead to ψ; however, this is not our

purpose. Instead, we assume that the problem of the isolated scatterer has been solved explicitly, and seek to solve a more complicated scattering problem in terms of this known solution ψ, or in terms of an auxiliary function which we shortly define.

Initially, we consider a cylinder which is symmetrical with respect to the yz plane, but which is otherwise arbitrary; see Fig. 1 for the geometry, and for definition of the co-ordinates used in the following. Later (Section V), we restrict discussion to cylinders with inversion symmetry (i.e., symmetrical to both yz and xz planes); finally (Section VI), we apply the results to the circular cylinder.
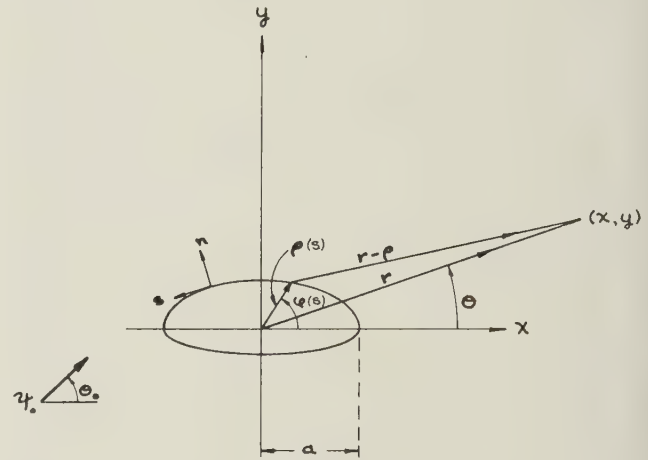
Fig. 1—Scattering of a plane wave by a cylinder symmetrical with respect to the y, z-plane. r, θ and $\rho(s)$, $\varphi(s)$ are polar co-ordinates of an observation point and a point on the surface respectively. n is the normal out of the surface and s is the co-ordinate around the perimeter. We also use $\mathbf{r} = \mathbf{x} + \mathbf{y}$, $\boldsymbol{\rho} = \boldsymbol{\xi} + \boldsymbol{\eta}$.

We apply Green's theorem to the pair of functions $u(\boldsymbol{\rho})$ and $G(\mathbf{r}, \boldsymbol{\rho})$, where $G = iH_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid)/4$ is the two dimensional free-space Green's function; i.e., $(\nabla^2 + k^2)G = -\delta(\mathbf{r} - \boldsymbol{\rho})$; H_o is the Hankel function of the first kind; and δ is the Dirac δ function. Integrating over the volume external to the scatterer, we express $u(\mathbf{r})$ as an integral over the scatterer's surface (the surface integral at infinity vanishing because of the radiation condition):

$$u(\mathbf{r}) = -\tfrac{1}{4}i \oint [H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid)\partial_n u(\boldsymbol{\rho})$$
$$- u(\boldsymbol{\rho})\partial_n H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid)] \, ds \equiv \{H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid), u(\boldsymbol{\rho})\}; \qquad (4)$$

the operation $\{ \quad \}$, here defined, represents a surface integral over one scatterer in all that follows. We find it convenient to add $\{H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid), \psi_o(\boldsymbol{\rho})\} = 0$ to (4); for future reference, we interpret the vanishing of this surface integral in terms of the vanishing of the equivalent *interior* volume integral—the volume integral vanishes since it contains no source of ψ_o. Thus we write

$$u(\mathbf{r}) = \{H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid), \Psi(\boldsymbol{\rho})\}. \qquad (4a)$$

If $kr \gg 1$, $r \gg \rho$, then

$$H_o(k \mid \mathbf{r} - \boldsymbol{\rho} \mid) \sim (2/i\pi kr)^{1/2}e^{ik[r - \rho \cos (\varphi - \theta)]}$$
$$\equiv H(kr)e^{-ik\rho \cos (\varphi - \theta)} \qquad (5)$$

The corresponding asymptotic form of u may thus be written

$$u(\mathbf{r}) \sim H(kr)g(\theta, \theta_o), \qquad (6)$$

$$g(\theta, \theta_o) \equiv \{e^{-ik\rho \cos(\varphi-\theta)}, \Psi(\boldsymbol{\rho}, \theta_o)\}. \qquad (7)$$

The scattering amplitude $g(\theta, \theta_o)$ indicates the far-field response in the direction θ to plane wave excitation of direction θ_o. As a consequence of the reciprocity theorem, we have in general $g(\theta, \theta_o) = g(\pi + \theta_o, \pi + \theta)$. In addition, because of the symmetry of the scatterer, we have

$$g(\theta, \theta_o) = g(\pi - \theta, \pi - \theta_o)$$
$$= \{e^{ik\rho \cos(\varphi+\theta)}, \Psi(\boldsymbol{\rho}, \pi - \theta_o)\}, \quad (8)$$

and also $\Psi(\rho(\varphi), \pi - \theta_o) = \Psi(\rho(\pi - \varphi), \theta_o)$.

It is also convenient to introduce the amplitudes

$$f_\pm (\theta, \pi - \theta_o) \equiv g(\theta, \theta_o) \pm g(\theta, \pi - \theta_o)$$
$$= f_\pm (\pi - \theta, \theta_o). \quad (9)$$

These amplitudes (twice the symmetric and antisymmetric components of g with respect to reflection of θ or θ_o in the plane of symmetry), are in themselves the scattering amplitudes of a boss (i.e., a protuberance) on a perfectly reflecting plane. The total wave functions for the boss problem,

$$\Psi(\mathbf{r}, \theta_o) \pm \Psi(\mathbf{r}, \pi - \theta_o) = \psi_o \pm \bar{\psi}_o + u_\pm = \Psi_\pm;$$
$$\psi_o = e^{ikr \cos(\theta+\theta_o)}; \quad (10)$$
$$u_\pm(\mathbf{r}, \pi - \theta_o) = u(\mathbf{r}, \theta_o) \pm u(\mathbf{r}, \pi - \theta_o)$$
$$\sim H(kr)f_\pm(\theta, \pi - \theta_o),$$

satisfy the same boundary conditions on the scatterer as Ψ; in addition, $\partial_x \Psi_+ = 0$, and $\Psi_- = 0$ on the plane $x = 0$. Of course, we may also construct f's such that f_+ and f_- correspond to different boundary conditions on the scatterer as well as on the plane; e.g., for a perfectly conducting boss on a perfectly conducting plane we require $\partial_n \Psi_+ = 0$ on both the scatterer and plane, for E-polarization perpendicular to the generator; and $\Psi_- = 0$, for parallel polarization. In the following, it will be clear from the context whether both f's satisfy the same or different conditions at the scatterer.

Although the scattering amplitude g is given in (7) in terms of the surface values of Ψ and $\partial_n \Psi$, it is in general simpler to obtain better approximations (or measurements) of the first function. More explicit representations than (7) exist for the elliptical cylinder, and the special case of the circular cylinder has received particular attention in the literature; however, representations of g suitable for practical purposes for many cases of physical interest are still to be discovered. Disregarding this, we assume in the following that g is known, and treat it as a "parameter" of the grating problem. By so doing, we can seek a formalism for the grating independent of its particular components, and concentrate on the novel features of the many-body scattering problem.

III. THE GRATING

We consider the two-dimensional problem of the scattering of a plane wave by a grating composed of an infinite number of parallel cylinders; see Fig. 2. The cylinders are identical to that of the previous section, and corresponding points of adjacent cylinders are separated a distance d along the y axis. We write the total field as

$$\Phi(\mathbf{r}) = \psi_o(\mathbf{r}) + U(\mathbf{r}), \qquad U(\mathbf{r}) = \sum_{s=-\infty}^{\infty} u_s(\mathbf{r} - s\mathbf{d}),$$
$$s = 0, \pm 1, \cdots, \quad (11)$$

where the total scattered wave U is expressed as a superposition of the waves scattered by the individual cylinders. The wave scattered by the sth cylinder is to fulfill the radiation condition that it be outgoing from s as $|\mathbf{r} - s\mathbf{d}| \to \infty$. The total field $\Phi(\mathbf{r})$ satisfies the same wave equation (as do its components) in the region exterior to all scatterers, and the same boundary conditions at all scatterers as Ψ of the previous section. The total scattered wave $U(\mathbf{r})$ is to fulfill the radiation condition that it consist of outgoing waves (here plane waves), and that it be bounded as $r \to \infty$; it will be shown that this is already insured by the radiation condition on u_s.
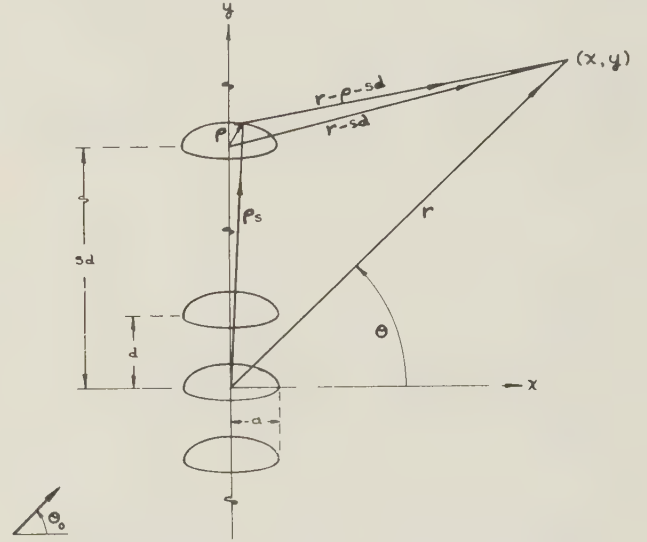
Fig. 2—Scattering of a plane wave by an infinite grating of identical cylinders. We use $|\mathbf{r} - \boldsymbol{\rho} - s\mathbf{d}|^2 = R_s^2 = (x - \xi)^2 + (y - \eta - sd)^2$, where $\xi = \rho \cos l$, $\eta = \rho \sin l$.

The radiation condition on $u_s(\mathbf{r} - s\mathbf{d})$ allows us to specify U without performing a limiting process, and with no more physical insight than that required by the problem of a single scatterer. Thus for the "central cylinder" (at whose origin the phase of the incident wave is zero), we apply Green's theorem to $u_o(\boldsymbol{\rho})$ and $G(\mathbf{r}, \boldsymbol{\rho})$. Integrating over the volume external to all scatterers, we obtain initially

$$u_o(\mathbf{r}) = \sum_s \{H_o(k|\mathbf{r} - \boldsymbol{\rho}_s|), u_o(\boldsymbol{\rho}_s)\}_s$$
$$= \{H_o(k|\mathbf{r} - \boldsymbol{\rho}|), u_o(\boldsymbol{\rho})\}, \quad (12)$$

where the final integration is over the surface of the central cylinder; the integral at infinity vanished be-

cause of the radiation condition on u_o; the integrals of u_o over the remaining cylinders vanished because the equivalent internal volume integrals contained no sources of u_o. Because of this last consideration we may also add $\{H_o(h \mid r - \rho \mid\mid), u_1(\rho)\} = 0$, etc., to $u_o(r)$; thus we may write

$$u_o(r) = \{H_o(h \mid r - \rho \mid), \Phi_o(\rho)\}, \qquad (12a)$$

where $\Phi_o(\rho)$ represents the *total* field at scatterer 0—all integrals we added to (12) to obtain (12a) vanish since ψ_o, u_1, etc., have no sources within surface 0. Similarly the wave scattered by the sth cylinder is obtained by replacing 0 by s, and r by $r - sd$. Thus, in general,

$$u_s(r - sd) = \{H_o(k \mid r - sd - \rho \mid\mid), \Phi_o(\rho)\}$$
$$= e^{iksd \sin \theta_0}\{H_o(k \mid r - sd - \rho \mid\mid), \Phi_s(\rho)\}, \quad (13)$$

where $\Phi_s(\rho) = e^{iksd \sin \theta_0} \Phi_o(\rho)$, the total field at the sth scatterer, differs from that at the zeroth only by the phase factor introduced by the incident wave.

Proceeding as for (6), we write the asymptotic form of u_o as

$$u_o(r) \sim H(kr)G(\theta, \theta_o), \qquad (14)$$

where

$$G(\theta, \theta_o) \equiv \{e^{-ik\rho \cos(\varphi-\theta)}, \Phi_o(\rho, \theta_o)\},$$
$$\rho \cos(\varphi - \theta) = \xi \cos\theta + \eta \sin\theta, \quad (15)$$

is the multiple scattered amplitude of a cylinder in the grating. Because of the symmetry of the grating, the present function G has the same symmetry property as the single scattered amplitude g of (7):

$$G(\theta, \theta_o) = G(\pi - \theta, \pi - \theta_o)$$
$$= \{e^{ik\rho \cos(\varphi+\theta)}, \Phi_o(\rho, \pi - \theta_o)\}. \quad (15a)$$

We show shortly that it is this function which characterizes the exact solution of the grating; then, we seek to specify G solely in terms of its single scattered value g. Note that $\Phi_o(\rho(\varphi), \pi - \theta_o) = \Phi_o(\rho(\pi - \varphi), \theta_o)$ is also symmetrical.

From (11) and (13) we write the total scattered wave as

$$U = \sum u_s = \sum e^{iksd \sin \theta_0}\{H_o(kR_s), \Phi_o(\rho)\}$$
$$\equiv \{\textstyle\sum, \Phi_o\}, \quad (16)$$

$$\textstyle\sum \equiv \sum_{s=-\infty}^{\infty} H_o(kR_s)e^{iksd \sin \theta_0};$$

$$R_s{}^2 = \mid r - sd - \rho \mid^2 = (x - \xi)^2 + (y - \eta - sd)^2, \quad (17)$$

where $(i/4)\,\Sigma$ is the Green's function for the reduced problem. The present representation of U is in terms of cylindrical waves. In order to obtain a plane wave representation we use[8]

$$H_o(k \mid r - \rho - sd \mid) = \frac{1}{\pi}\int_{-\infty}^{\infty} (1 - t^2)^{-1/2}$$

$$\exp[ik(y - \eta - sd)t + ik \mid x - \xi \mid (1 - t^2)^{1/2}]\,dt, \quad (18)$$

where $(1 - t^2)^{1/2} = i(t^2 - 1)^{1/2}$ for $t > 1$; this insures that the integral converges. Substituting (18) into (17), we note that the sum over s equals the periodic Dirac δ function

$$\sum_{s=-\infty}^{\infty} e^{-iksd(t - \sin\theta_0)} = \frac{2\pi}{kd}\sum_{v=-\infty}^{\infty}\delta\left(t - \sin\theta_o - \frac{2v\pi}{kd}\right),$$
$$v = 0, \pm 1, \cdots. \quad (19)$$

Integrating over t in the new form of (17), we obtain the known result (*MO*, p 62)

$$\textstyle\sum = 2\sum_{v=-\infty}^{\infty} C_v e^{ik(y-\eta)\sin\theta_v + ik|x-\xi|\cos\theta_v},$$
$$C_v = (kd \cos\theta_v)^{-1}, \quad (20)$$

where

$$\sin\theta_v \equiv \sin\theta_o + 2\pi v/kd = \sin\theta_o + v\lambda/d,$$
$$\cos\theta_v = (1 - \sin^2\theta_v)^{1/2}. \quad (21)$$

The angles θ_v are the usual "diffraction angles" of the grating. The "propagating modes" correspond to $\mid \sin\theta_v \mid \leqslant 1$, or equivalently to $v_- \leqslant v \leqslant v_+$, where $\mid v_\pm \mid \leqslant (1 \pm \sin\theta_o)\,d/\lambda < \mid v_\pm + 1 \mid$. The nonpropagating modes (the "surface waves") correspond to the remaining values of v; for these modes we have $\cos\theta_v = i(\sin^2\theta_v - 1)^{1/2}$, in accord with the convergence requirement on the integral in (18); this form of $\cos\theta_v$, which insures fulfillment of the radiation condition on U, is thus fixed by the radiation condition on $u_s(r - sd)$.

Substituting (20) into (16), we obtain

$$U(r, \theta_o) = 2\Sigma C_v\{e^{ik(y-\eta)\sin\theta_v + ik|x-\xi|\cos\theta_v}, \Phi_o(\rho, \theta_o)\}. \quad (22)$$

Hence if $x > \xi_{max} \equiv a$, then from definition (15) it follows that

$$\Phi = \psi_o + U = \psi_o + 2\Sigma C_v\psi_v G(\theta_v, \theta_o),$$
$$\psi_v = e^{ikr \cos(\theta - \theta_v)}, x > a. \quad (23)$$

Similarly if $x < -a$, then

$$\Phi = \psi_o + 2\Sigma C_v\bar{\psi}_v G(\pi - \theta_v, \theta_o),$$
$$\bar{\psi}_v = e^{-ikr \cos(\theta + \theta_v)}, x < -a. \quad (24)$$

If $\mid x \mid \to \infty$, then only the propagating modes persist; here Φ of (23) is the total field transmitted by the grating, and $\Phi - \psi_o$ of (24) is the total reflected field. For $\mid x \mid \to \infty$ these expressions could have been written down directly from the physical requirement that $\Phi e^{-iky \sin\theta_0}$ have the period d in y, and from elementary considerations based on Fresnel zones; note that $2\sqrt{\lambda r}\sec\theta_v$ is the width of the first Fresnel zone (here two strips) of mode v, and d^{-1} is the number of scatterers in unit length of grating—each propagating mode is proportional to the wave scattered by the central element of the zone appropriate to v, and to half the number of elements in the zone.

The above plane wave representations differ from those of previous writers[6,7] in the way the amplitudes have been expressed. Thus the usual "transmission amplitudes" of modes 0, v, etc., equal our 1 +

---

[8] W. Magnus and F. Oberhettinger, "Formeln und Satze fur die Speziellen Funktionen der Mathematischen Physik," Springer, Berlin; p. 40; 1948. This text is hereafter referred to as MO.

$2C_oG(\theta_o, \theta_o)$, $2C_vG(\theta_v, \theta_o)$, etc. The present resolution of these amplitudes isolates the incident part of the field from the scattered part, and regards the scattered part as made up of a "Fresnel factor" $2C_v$ times the multiple scattered amplitude of one cylinder. The utility of this will be indicated in the following.

On the other hand, for $-a < x < a$ ("inside the grating", but not within a scatterer), we may write

$$\Phi = \psi_o + U = \psi_o + 2\Sigma C_v[\psi_v G_L(\theta_v, \theta_o) + \bar{\psi}_v G_R(\pi - \theta_v, \theta_o)], |x| \leqslant a, \quad (25)$$

where G_L and G_R are "incomplete scattering amplitudes", L and R indicating paths of integration corresponding to $x > \xi$ and $x < \xi$ respectively. Thus, in this region, the scattered field at x consists of the forward modes of the left part of the grating, and the back modes of the right part. (Lipmann[6] was the first to obtain a similar representation for the field within the valleys of a corrugated reflection grating.)

We now seek a more explicit representation for the G's than given by (15). We add and subtract the wave scattered by the central element to the total field at its surface:

$$\Phi_o = \psi_o + U = \psi_o + (U - u_o) + u_o, \quad (25a)$$

where U is given by (22), and u_o by (12a). We first rewrite $U - u_o$ wholly in terms of plane waves; the resulting expression, which may be regarded as the solution for an isolated scatterer excited by a superposition of plane waves, may be "solved" for the G's by inspection.

We rewrite $H_o(k|\mathbf{r} - \boldsymbol{\rho}|)$ of (18) as the integral analogous to the sum of (20). Thus letting $t = \sin\theta_\mu = \sin\theta_o + 2\pi\mu/kd$, we obtain

$$H_o(k|\mathbf{r} - \boldsymbol{\rho}|) = 2\int_{-\infty}^{\infty} C_\mu e^{ik(y-\eta)\sin\theta_\mu + ik|x-\xi|\cos\theta_\mu} d\mu,$$
$$C_\mu = (kd\cos\theta_\mu)^{-1}. \quad (26)$$

Substituting (26) into u_o of (12a), and then subtracting the result from (22), we write $U - u_o$ as the difference of the discrete and continuous spectra:

$$U - u_o = S2C_v\{e^{ik(y-\eta)\sin\theta_v + ik|x-\xi|\cos\theta_v}, \Phi_o(\boldsymbol{\rho}(\varphi), \theta_o)\}$$
$$= S2C_v[\psi_v G_L(\theta_v, \theta_o) + \bar{\psi}_v G_R(\pi - \theta_v, \theta_o)], |x| \leqslant a, \quad (27)$$

where the operator S is defined as the difference of the sum and integral operations,

$$S \equiv \sum_{v=-\infty}^{\infty} - \int_{-\infty}^{\infty} d\mu = \int_{-\infty}^{\infty} d\mu[\sum \delta(\mu - v) - 1]. \quad (28)$$

It is clear that contributions to $U - u_o$ vanish asymptotically with v and μ.

We now use the restriction that the scatterers are symmetrical with respect to the yz plane. Because of this, the grating itself is symmetrical, and consequently (27) must be unaltered if we reflect both the directions of incidence and observation. Thus replacing θ and θ_o by their supplements (so that ψ_v becomes $\bar{\psi}_v$, etc.), we rewrite (27) as

$$U - u_o = S2C_v[\psi_v G_{R'}(\theta_v, \pi - \theta_o) + \bar{\psi}_v G_{L'}(\pi - \theta_v, \pi - \theta_o)], \quad |x| \leqslant a, \quad (29)$$

where R' and L' are the images of the previous paths L and R respectively. We now reflect the integration variable (i.e., replace ξ by $-\xi$, or φ by $\pi - \varphi$) in the integrals involved in the G's, and use the symmetry property $\Phi_o(\pi - \varphi, \pi - \theta_o) = \Phi(\varphi, \theta_o)$; thus we obtain

$$U - u_o = S2C_v[\psi_v G_L(\pi - \theta_v, \theta_o) + \psi_v G_R(\theta_v, \theta_o)],$$
$$|x| \leqslant a. \quad (29a)$$

Expressing $U - u_o$ as the mean of the two forms (27) and (29a), we obtain finally

$$U - u_o = SC_v[\psi_v G(\theta_v, \theta_o) + \bar{\psi}_v G(\pi - \theta_v, \theta_o)],$$
$$|x| \leqslant a, \quad (30)$$

where the present kernel is also the mean of the kernels of (23) and (24), i.e., of the forms on opposite sides of the grating.[9] Although both U and u_o may diverge individually with v or $\mu \rightarrow \infty$, their difference remains finite.

The field at the central scatterer may therefore be written

$$\Phi_o = \psi_o + SC_v[\psi_v G(\theta_v, \theta_o) + \bar{\psi}_v G(\pi - \theta_v, \theta_o)] + u_o; \quad (31)$$

it consists of the original plane wave, the difference of sets of discrete and continuous plane waves, and the outgoing scattered wave. Since this form is identically that for an isolated scatterer excited by a superposition of plane waves, the corresponding scattering amplitude of u_o is given by a superposition of the scattering amplitudes of the single cylinder. Thus by comparison of (31) with (2) and (7), we write directly

$$G(\theta, \theta_o) = g(\theta, \theta_o)$$
$$+ SC_v[g(\theta, \theta_v)G(\theta_v, \theta_o)$$
$$+ g(\theta, \pi - \theta_v)G(\pi - \theta_v, \theta_o)]. \quad (32)$$

The leading term of the amplitude is the corresponding single scattered value; the remainder corresponds to the effects of multiple scattering. We have thus represented G as the solution of an inhomogeneous "sum-integral" equation.

The symmetric and antisymmetric components of G with respect to reflection of one angle,

$$F_\pm(\theta, \pi - \theta_o) \equiv G(\theta, \theta_o) \pm G(\theta, \pi - \theta_o)$$
$$= F_\pm(\pi - \theta, \theta_o), \quad (33)$$

may be expressed in terms of an analogous relation involving their single scattered values $f_\pm$ as defined in (9). Substituting into (32) we obtain, for either F_+ or F_-, the somewhat simpler relation

---

[9] It should be noted that the reflection procedure leading to (29a) is equivalent to replacing θ_v by $\pi - \theta_v$ in $\{\}$ of (27). This replacement could perhaps have been done directly, since the definition of $\sin\theta_v$ in (21) applies equally well for $\sin(\pi - \theta_v)$; this arbitrariness was resolved for $|x| > a$ by the radiation condition, a condition not necessarily limiting the field for $|x| < a$. Thus in distinction to the "outgoing" form (27), we may interpret (29a) as the "incoming" form, and regard (30) as the "standing wave" form—much in the same fashion as we interpret $H_o^{(1)} + H_o^{(2)} = 2J_o$ inside a cylinder. If this consideration is valid, then (30) is not necessarily restricted to symmetrical scatterers.

$$F(\theta, \pi - \theta_o) = f(\theta, \pi - \theta_o)$$
$$+ SC_v f(\theta, \pi - \theta_v) F(\pi - \theta_v, \theta_o). \quad (33a)$$

The F's are in themselves the appropriate scattering amplitudes of a boss in the reflection grating (bosses on a perfectly conducting plane) of Fig. 3. The total field for this grating is given by

$$\Phi_\pm = \pm \psi_o + \Phi_{r\pm},$$
$$\Phi_{r\pm} = \psi_o + 2 \sum \psi_v C_v F_\pm(\theta_v, \pi - \theta_o), \quad x > a, \quad (34)$$

where the original excitation is taken as $\pm \bar{\psi}_o$ for the two cases. (For electromagnetics, the two cases plus and minus correspond respectively to E polarization perpendicular and parallel to the generator of the reflection grating; for this application, the two f's satisfy different boundary conditions at the single cylinder as determined by the polarization. If the bosses are perfectly conducting, then the present grating is equivalent to the perfectly conducting corrugated grating.[6])
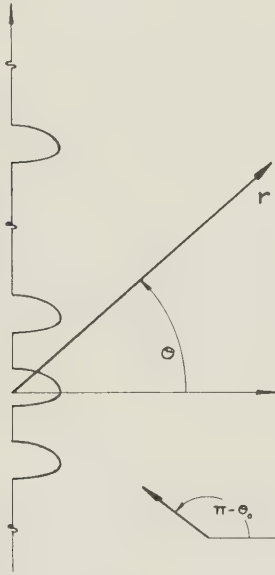
Fig. 3—Scattering of a plane wave by a reflection grating of bosses (protuberances) on a perfectly reflecting plane.

Before discussing the relations (32) and (33) for G and F further, we derive certain theorems which the amplitudes fulfill; these are used to motivate the analysis of Section V, and to check the approximations.

IV. Scattering Theorems

The various relations we consider in this section follow from one theorem; however, we initially derive certain special cases by means of the energy principle because of the physical insight this leads to. For simplicity, this section is restricted to nonabsorbing scatterers; the material may be extended by introducing appropriate "absorption cross sections."

The total time-averaged energy flux (normalized through division by the flux density of the incident wave) may be written as $\mathbf{I} = \text{Re}\,(\Phi^* \nabla \Phi / ik)$. If the cylinders are nonabsorbing, then $\nabla \cdot \mathbf{I} = 0$ everywhere, and consequently $\int \nabla \cdot \mathbf{I}\, d\tau = \int \mathbf{I} \cdot d\mathbf{s} = 0$; i.e., there is no

average energy outflow from any enclosed volume. We choose the volume bounded by $x = \pm \infty$ and $y = \pm d/2$. The integrals over the faces $y = \pm d/2$ cancel because of the periodicity of $\psi_v \psi_\mu^*$, and because of the oppositely directed surface normals; the periodicity also eliminates the interference terms in the integrals along y at $x = \pm \infty$. Using the two forms (23) and (24) for $x \to \pm \infty$, and indicating the corresponding scattered modes by V_v and $\bar{V}_v$, we therefore obtain

$$\psi_o^* \partial_x \psi_o / ik + \sum V_v^* \partial_x V_v / ik$$
$$= (\psi_o + V_o)^* \partial_x \chi(\psi_o + V_o)/ik + \sum{}' V_v^* \partial_x V_v / ik, \quad (35)$$

where only the propagating modes exist in the sums, and where $\sum'$ indicates $v \neq 0$. We substitute

$$V_v = 2 C_v \psi_v G(\theta_v, \theta_o), \quad C_v = (k\, d \cos \theta_v)^{-1};$$

thus, since $\partial_x \psi_v / ik = \psi_v \cos \theta_v$, etc., we obtain

$$- \text{Re}\, G(\theta_o, \theta_o) = \sum C_v [\,|\, G(\theta_v, \theta_o)\,|^2.$$
$$+ |\, G(\pi - \theta_v, \theta_o)\,|^2\,], \qquad \sum = \sum_{v=v_-}^{v_+}. \quad (36)$$

This theorem states that the energy flux lost from the incident wave by interference with the central transmitted mode of the grating appears in the transmitted and reflected modes.

Applying the same procedure to the reflection grating [here the volume is bounded by $x = 0$ (where $\mathbf{I} \cdot \mathbf{x} = 0$), and $_x = \infty$] we obtain the analogous theorem

$$- \text{Re}\, F(\theta_o, \pi - \theta_o) = \sum C_v |\, F(\theta_v, \pi - \theta_o)\,|^2. \quad (37)$$

If the cylinders are symmetrical with respect to the yz plane, then the definition of F together with theorems (36) and (37) lead to

$$- \text{Re}\, G(\theta_o, \pi - \theta_o)$$
$$= 2\, \text{Re} \sum C_v G^*(\theta_v, \theta_o) G(\theta_v, \pi - \theta_o). \quad (38)$$

For comparison, we note the corresponding theorems for an isolated cylinder[10]

$$- \text{Re}\, g(\theta_o, \theta_o) = \frac{1}{2\pi} \int_o^{2\pi} |\, g(\theta, \theta_o)\,|^2\, d\theta$$

$$= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |\, g(\theta, \theta_o)\,|^2 + |\, g(\pi - \theta, \theta_o)\,|^2\,]\, d\theta, \quad (36a)$$

$$- \text{Re}\, f(\theta_o, \pi - \theta_o)$$
$$= \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |\, f(\theta, \pi - \theta_o)\,|^2\, d\theta, \quad (37a)$$

$$- \text{Re}\, g(\theta_o, \pi - \theta_o)$$
$$= \text{Re}\, \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} g(\theta, \theta_o) g^*(\theta, \pi - \theta_o)\, d\theta; \quad (38a)$$

the first (or second) expression multiplied by $4/k$ equals the total scattering cross section per unit length of cylinder (or boss); i.e., the power scattered out of the incident wave by unit length of cylinder (or boss). (Our definition of g differs from the usual one. We defined g by dividing the far field scattered wave by the asymptotic value of $H_o(kr)$ rather than by $r^{1/2} e^{ikr}$ as is

[10] V. Twersky, *Jour. Appl. Phys.* vol. 25, pp. 859; 1954.

customary; our motivation is that the present definition allows us to interpret the right side of (36a) as the average of $|g|^2$ over all values of θ; similarly (36) may be interpreted as the corresponding average over the discrete propagating values.)

We may rewrite (36a) etc., in terms of $\mu = kd \,(\sin \theta_\mu - \sin \theta_o)/2\pi$:

$$- \operatorname{Re} g\,(\theta_o, \theta_o)$$
$$= \int_{u-}^{\mu_+} C_\mu [\,|\,g(\theta_\mu, \theta_o >\,|^2 + |\,g(\pi - \theta_\mu, \theta_o)\,|^2\,]\,d\mu,$$

$$\mu_\pm = kd(\pm 1 - \sin \theta_o)/2\pi, \quad (36b)$$

etc. The essential difference between the unprimed and primed equations is thus that the former indicates scattering in certain discrete directions, and the latter a continuum. Since the number of propagating modes is of the order of kd, it is clear physically how the result for the grating goes over to that of the isolated scatterer as $kd \to \infty$.

In our representation, the central transmitted mode of the grating is written as the sum of the incident and scattered components. We recast our results in a more familiar form by introducing the "transmission" and "reflection amplitudes"

$$t_v \equiv \delta_{vo} + 2C_v G(\theta_v, \theta_o)$$
$$r_v = 2C_v G(\pi - \theta_v, \theta_o), \quad (39)$$

where δ_{vo} is the Kronecker δ. In terms of these amplitudes, we write the transmitted and reflected fields as

$$\Phi_t = \psi_i + U = \sum t_v \psi_v, \; x > a;$$
$$\Phi_r = U = \sum r_v \bar{\psi}_v, \; x < -a. \quad (40)$$

We also rewrite theorems (36) and (38) as

$$\cos \theta_o = \sum \,(\,|\,t_v\,|^2 + |\,r_v\,|^2)\,\cos \theta_v$$
$$\equiv \mathbf{I}_o \cdot \mathbf{i} = \sum \,(\mathbf{T}_v - \mathbf{R}_v) \cdot \mathbf{i}, \quad (41)$$

$$\operatorname{Re} \sum t_v{}^* r_v \cos \theta_v = 0, \quad (42)$$

where

$$\mathbf{T}_v \equiv |\,t_v\,|^2 \,\mathbf{v}, \qquad \mathbf{R}_v \equiv |\,r_v\,|^2 \,\bar{\mathbf{v}};$$
$$\mathbf{v}, \mathbf{v} = \pm \mathbf{i} \cos \theta_v + \mathbf{j} \sin \theta_v, \quad (43)$$

are the normalized Poynting (or acoustic) intensities of the individual modes, and $|\,\mathbf{I}_o\,|^2 = 1$ is the incident value; (41) thus expresses continuity of the summed modal components of energy flux normal to the plane of the grating. The utility of the "no coupling" relation (42) may be illustrated for the case of one propagating mode: thus (41) states $|\,t\,|^2 + |\,r\,|^2 = 1$; adding the result $\operatorname{Re} t^* r = 0$ of (42) gives $|\,t + r\,| = 1$, which restricts the amplitudes instead of merely the intensities.

Similarly for the reflection grating we introduce

$$\rho_v = \delta_{vo} + 2C_v F(\theta_v, \pi - \theta_o) = t_v \pm r_v, \quad (44)$$

and write

$$\Phi_r = \sum \rho_v \psi_v, \quad (45)$$

$$\cos \theta_o = \sum |\,\rho_v\,|^2 \cos \theta_v \equiv -\mathbf{I}_o \cdot \mathbf{i} = \sum \mathbf{R}_v \cdot \mathbf{i},$$
$$\mathbf{R}_v \equiv |\,\rho_v\,|^2 \mathbf{v}. \quad (46)$$

The above may be extended to take into account absorption of energy by the cylinders (i.e., dissipation as heat), by adding the absorption cross section of one cylinder in the grating to the right sides of (35) and (36), and by adding the analogous function to the right side of (37). These functions, say P for (36), may be specified in terms of their single scattered analogs together with g and G; for our present purposes, however, we may simply add P to the right side of (36) in order to obtain its analytical representation.

We now derive a theorem which includes the previous as special cases. We rewrite the previous $\Phi(\theta_o)$ as Φ_o, and similarly we write $\Phi(\theta_v)$ as Φ_v, where the subscript indicates the direction of arrival of the incident wave. We have

$$\Phi_v = \psi_v + 2 \sum_\alpha C_\alpha G(\theta_\alpha, \theta_v)\psi_\alpha, \qquad x > a, \quad (47)$$

and the corresponding expression containing $G(\pi - \theta_\alpha, \theta_v)\psi_\alpha$ for $x < -a$; also $(\nabla^2 + k^2)\Phi_v = 0$ in the region external to the scatterers.

We apply Green's theorem to the pair of functions $\Phi_\mu{}^*$ and Φ_v in the volume bounded on the outside by $x = \pm \infty$, $y = \pm d/2$, and on the inside by the surface of the inclosed cylinder. Again, we assume that there is no absorption; i.e., that either Φ or $\partial_n \Phi$ vanishes at the surface of the scatterer (here a perfect conductor), or else that $(\nabla^2 + h^2)\Phi = 0$, $\operatorname{Im} h = 0$ inside the scatterer (here a perfect dielectric). Consequently

$$\int \,[\Phi_\mu{}^* \nabla^2 \Phi_v - \Phi_v \nabla^2 \Phi_\mu{}^*]\,d\tau = 0 \to \oint \,[\Phi_\mu{}^* \nabla(\Phi_v/ik)$$
$$+ \Phi_v \nabla(\Phi_\mu/ik)^*]\cdot d\mathbf{s} = 0; \quad x = \pm \infty, \; y = \pm \frac{d}{2}. \quad (48)$$

The integration is performed as previously (i.e., the only functions of the integration variables are again $\psi_v{}^* \psi_\mu$), and we obtain

$$-G(\theta_\mu, \theta_v) - G^*(\theta_v, \theta_\mu) = 2 \sum_\alpha C_\alpha [G^*(\theta_\alpha, \theta_\mu)G(\theta_\alpha, \theta_v)$$
$$+ G^*(\pi - \theta_\alpha, \theta_\mu)G(\pi - \theta_\alpha, \theta_v)]. \quad (49)$$

If we allowed absorption, then the surface integral in (48) would also extend over the surface of one cylinder, and the right side of (49) would contain a corresponding function $P_{v\mu}$; the special value P_{vv} corresponds to the absorption cross section of a cylinder in the grating for the angle of incidence θ_v.

From (49) we obtain the previous relations: if $\theta_v = \theta_\mu = \theta_o$, the result reduces to (36); if $\theta_v = \pi - \theta_\mu = \theta_o$, and if the amplitudes are symmetrical to reflection, we obtain (38); these two special cases lead to (37). We may also obtain (37) more directly by specializing the reflection theorem analogous to (49):

$$-F(\theta_\mu, \pi - \theta_v) - F^*(\theta_v, \pi - \theta_\mu)$$
$$= 2 \sum_\alpha C_\alpha F^*(\theta_\alpha, \pi - \theta_\mu)F(\theta_\alpha, \pi - \theta_v); \quad (50)$$

this result decomposes into two forms of (49).

The theorem analogous to (49) for an isolated cylinder may be written as[11]

$$-g(\theta_\mu, \theta_v) - g^*(\theta_v, \theta_\mu) = \frac{1}{\pi} \int_0^{2\pi} g^*(\theta, \theta_\mu) g(\theta, \theta_v) \, d\theta$$

$$= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} [g^*(\theta, \theta_\mu) g(\theta, \theta_v)$$

$$+ g^*(\pi - \theta, \theta_\mu) g(\pi - \theta, \theta_v)] \, d\theta; \quad (49)$$

this reduces to $-2 \, \mathrm{Re} \, g(\theta_\mu, \theta_v)$ if the cylinders have inversion symmetry. Similarly the analog of (50) is

$$-f(\theta_\mu, \pi - \theta_v) - f^*(\theta_v, \pi - \theta_\mu)$$

$$= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} f^*(\theta, \pi - \theta_\mu) f(\theta, \pi - \theta_v) \, d\theta, \quad (50a)$$

which reduces to $-2 \, \mathrm{Re} \, f(\theta_\mu, \pi - \theta_v)$ for symmetrical bosses. (This last relation applies to a boss on a perfectly reflecting plane; the special form (37a) relates the scattering cross section of the boss to its scattering amplitude in the direction of specular reflection of the plane. An analogous theorem for a scatterer at the intersection of two right-angled planes may be obtained from the above by replacing $\pi - \theta_v$ in (50a) by $\pi + \theta_v$, etc., and by replacing the lower limit of the integral by 0; the appropriate amplitudes are superpositions of four imaged g's. A special case of this theorem relates the scattering cross section to the back scattered amplitude.)

Finally, we may rewrite (49) and (50) in terms of amplitudes analogous to those of (39) and (44) as

$$\sum_\alpha (t_{\alpha\mu}{}^* t_{\alpha v} + r_{\alpha\mu}{}^* r_{\alpha v}) \cos \theta_\alpha = 0; \quad t_{\alpha\beta} = \delta_{\alpha\beta}$$
$$+ 2C_\alpha G(\theta_\alpha, \theta_\beta), \quad r_{\alpha\beta} = 2C_\alpha G(\pi - \theta_\alpha, \theta_\beta). \quad (51)$$

$$\sum_\alpha \rho_{\alpha\mu}{}^* \rho_{\alpha v} \cos \theta_\alpha = 0;$$

$$\rho_{\alpha\beta} = \delta_{\alpha\beta} + 2C_\alpha F(\theta_\alpha, \pi - \theta_\beta). \quad (52)$$

If there are n propagating modes, or equivalently n^2 values of $\rho_{\alpha\beta}$ of interest, then (52) gives n^2 relations between n pairs of ρ's.

V. The Scattering Amplitudes

We now consider further the expression for the scattering amplitude derived in Section III. We split the range of the operator $S = \Sigma_v - \int d\mu$ of (32) in two. We write $S = S_p + S_s$, where the first corresponds to the propagating waves, and the second to the surface waves. Noting that the complete integral operation involved may be written

$$\int_{-\infty}^\infty C_\mu [\quad] \, d\mu = \int_{-\pi/2+i\infty}^{\pi/2-i\infty} [\quad] \frac{d\theta_\mu}{2\pi},$$

and that the propagating part equals

$$\frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} [g(\theta, \varphi) G(\varphi, \theta_o) + g(\theta, \pi - \varphi) G(\pi - \varphi, \theta_o)] \, d\varphi$$

$$= \frac{1}{2\pi} \int_0^{2\pi} g(\theta, \varphi) G(\varphi, \theta_o) \, d\varphi, \quad (53)$$

we rewrite (32) initially as

$$G(\theta, \theta_o) = g(\theta, \theta_o) + L(g, G)$$

$$- \frac{1}{2\pi} \int_0^{\pi 2} g(\theta, \varphi) G(\varphi, \theta_o) \, d\varphi,$$

$$L(g, G)$$
$$= \overline{SC}_v [g(\theta, \theta_v) G(\theta_v, \theta_o) + g(\theta, \pi - \theta_v) G(\pi - \theta_v, \theta_o)],$$
$$\overline{S} = \sum_p + S_s. \quad (54)$$

We introduce the "transformed single scattering amplitudes"

$$\bar{g}(\theta, \theta_o) = g(\theta, \theta_o) - \frac{1}{2\pi} \int_0^{2\pi} g(\theta, \varphi) \breve{g}(\varphi, \theta_o) \, d\varphi, \quad (55)$$

which we rewrite operationally as

$$\frac{1}{2\pi} \int_0^{2\pi} [g(\theta, \varphi) + 2\pi \delta(\varphi - \theta)] \bar{g}(\varphi, \theta_o) \, d\varphi = g(\theta, \theta_o)$$

$$\to P\bar{g} = g, \quad \bar{g} = P^{-1}g, \quad (55a)$$

where, by P^{-1} (whose existence we assume) we mean the inverse of the integral operation on the left. Similarly we may write (54) operationally as

$$PG = g + L(g, G),$$
$$G = P^{-1}[g + L(g, G)] = \bar{g} + L(\bar{g}, G). \quad (54a)$$

Consequently, we obtain

$$G(\theta, \theta_o) = \bar{g}(\theta, \theta_o) + \overline{SC}_v [\bar{g}(\theta, \theta_v) G(\theta_v, \theta_o)$$

$$+ \bar{g}(\theta, \pi - \theta_v) G(\pi - \theta_v, \theta_o)], \quad \overline{S} = \sum_p + S_s, \quad (56)$$

where S excludes the continuous "propagating range." The original form may be recovered by dropping the bars; however, the transformed amplitudes are more convenient to work with. We illustrate this in the following.

We restrict discussion to cylinders with inversion symmetry, and show that for this case $g(\theta, \theta_o)$ is imaginary, provided the cylinder is nonabsorbing, and the angles real. Thus we write

$$\bar{g} = P^{-1}g = P^{-1}[(P^{-1})^* P^*]g = | P^{-1} |^2 [g(\theta, \theta_o)$$

$$+ \frac{1}{2\pi} \int_0^{2\pi} g^*(\theta, \varphi) g(\varphi, \theta_o) \, d\varphi]. \quad (57)$$

Hence, since we assumed inversion symmetry, we may use theorem (49) for $g(\varphi, \theta) = g(\theta, \varphi)$. Consequently, the integral in (57) equals $-\mathrm{Re} \, g(\theta, \theta_o)$, and

$$\bar{g}(\theta, \theta_o) = | p^{-1} |^2 i \, \mathrm{Im} \, g(\theta, \theta_o). \quad (58)$$

The fact that $\bar{g}$ is imaginary for this case will facilitate application of the energy theorems of the preceding section. For simplicity we also restrict the following discussion to nonabsorbing scatterers.

We may represent the nonpropagating part of (56) as a series by means of the Euler summation formula[12]: thus

$$S_s M_v = (\Sigma - \int d\mu)_s M_v$$

$$= - \left[\int_{\pi/2+i\gamma_+}^{\pi/2-io} + \int_{-\pi/2+io}^{-\pi/2+i\gamma_-} \right] \frac{M_\mu}{C_\mu} \, d\theta_\mu$$

$$+ \left[\tfrac{1}{2} M_m - \frac{B_1}{2!} M_{m'} + \frac{B_2}{4!} M_{m'''} \cdots \right]_{m = v_\pm \pm 1} \quad (59)$$

[11] R. Glauber and V. Schomaker, *Phys. Rev.*, vol. 84, p. 667; 1953.

[12] G. H. Hardy, "Divergent Series." Oxford University Press, Oxford, p. 318; 1949.

where M_ν represents the kernel of (56); $\nu_\pm \pm 1$ are the first nonpropagating modes, and $\pm(\pi/2 - i\gamma_\pm)$ are the corresponding values of θ; the B's are the Bernoullian polynomials. (We have ignored terms at ∞.) Successive terms of the series involve higher powers of $C_m = (h\,d\,\cos\theta_m)^{-1}$, and consequently the surface terms are small provided that $h\,d \gg 1$, $\gamma_\pm \to 0$. On the other hand, the above indicates that difficulties may arise if, e.g., $\pi/2 - i\gamma_+ = \pi/2 - i\epsilon \to \pi/2$, $C_m \to \infty$. However, although we expect interesting effects in this range,[4] the behavior at $\epsilon = 0$ must be representable by $\pi/2 - \epsilon \to \pi/2$ (i.e., the limit from the propagating side of the spectrum), for which case $C_m \approx (4\pi h\,d)^{-1/2}$, $m = \nu_+ + 1$, is small, and the surface waves negligible. Thus, since we expect essentially the same formal behavior for either limit ($\pi/2 - \epsilon$, $\pi/2 - i\epsilon$; $\epsilon \to 0$), we may heuristically approximate (59) by the leading nonpropagating terms for certain applications, or even neglect them completely for others; more generally we retain all of S_ϵ. (The special case of isotropic scatterers is treated further in Appendix A.)

Rather than work directly with (56), we find it simpler to discuss its components, i.e., the analogous relation for the reflection grating:

$$F(\theta,\ \pi - \theta_o) = \bar{f}(\theta,\ \pi - \theta_\nu)$$
$$+\ \bar{S}C_\nu\bar{f}(\theta,\ \pi - \theta_\nu)F(\pi - \theta_\nu,\ \theta_o),$$
$$F_\pm(\theta,\ \pi - \theta_o) = G(\theta,\ \theta_o) \pm G(\theta,\ \pi - \theta_o);$$
$$\bar{f}_\pm(\theta,\ \pi - \theta_o) = \bar{g}(\theta,\ \theta_o) \pm \bar{g}(\theta,\ \pi - \theta_o)$$
$$= f_\pm(\theta,\ \pi - \theta_o) - \tfrac{1}{2}\pi \int_{-\pi/2}^{\pi/2} f_\pm(\theta,\ \pi - \varphi)\bar{f}_\pm(\varphi,\ \pi - \theta_o)\,d\varphi,$$
$$\tag{60}$$

where $\bar{f}$ is regarded as known; here we used the fact that g has the same symmetry properties as $\bar{g}$ (as follows from (53) and (55)).

For brevity, we rewrite (60) as

$$F_\mu = \bar{f}_{\mu o} + \bar{S}C_\nu\bar{f}_{\mu\nu}F_\nu, \qquad \bar{f}_{\mu\nu} = \bar{f}_{\nu\mu}, \tag{61}$$

where the second relation holds for bosses symmetrical to the xz plane; because of the symmetry restriction and the restriction to nonabsorbing scatterers, f is imaginary, at least for real angles. The corresponding expression for the reflection amplitudes is

$$\rho_\mu = \delta_{\mu o} + 2C_\mu F_\mu = \delta_{\mu o} + C_\mu \bar{f}_{\mu o} + \bar{S}C_\mu \bar{f}_{\mu\nu}\rho_\nu, \tag{62}$$

or, equivalently,

$$\rho_\mu = \delta_{\mu o} + z_{\mu o} + \bar{S}z_{\mu\nu}\rho_\nu,$$
$$z_{\mu\nu} \equiv C_\mu \bar{f}_{\mu\nu} = z_{\nu\mu}\frac{\cos\theta_\nu}{\cos\theta_\mu}, \tag{62}$$

where the z's may be regarded as "impedances." The corresponding transmission and reflection amplitudes of the transmission grating are given by

$$t_\nu = \tfrac{1}{2}(\rho_\nu^+ + \rho_\nu^-), \qquad r_\nu = \tfrac{1}{2}(\rho_\nu^+ - \rho_\nu^-). \tag{63}$$

We consider first heuristically the special cases when only one and two modes are retained; then we consider the general case and the Wood anomalies. (One of our purposes in treating these special cases is to deduce certain additional properties of the f's.)

If we keep only one mode in (62) (i.e., if

$$d(1 \pm \sin\theta_o)/\lambda < 1,$$

and if we neglect the surface waves), then we obtain simply

$$\rho = 1 + z + z\rho = \frac{1 + z}{1 - z},$$
$$z = C\sigma\bar{f}_{oo} = \frac{\bar{f}(\theta,\ \pi - \theta)}{k\,d\,\cos\theta}; \tag{64}$$

for this case the grating is equivalent to a plane characterized by a "normal impedance" proportional to z (i.e., $\Phi/\partial_x\Phi$ is a constant at $x = 0$). Since z is imaginary, $|\rho|^2 = 1$ as required by the energy principle. If z is small, then $\rho \approx 1 + 2z + 2z^2$, where the term in z^2 must be retained in order to fulfill the energy theorem.

In order to consider the range $\theta \to \pi/2$ (grazing incidence), we must first specify the limiting behavior of $f_\pm$. It is clear physically that f_- and f_+ (the amplitudes of an isolated scatterer) behave quite differently for this range; the essential difference is that, as $\theta = \pi/2 - \epsilon \to \pi/2$, the plane wave excitation vanishes like ϵ^2 for the first, and is nonvanishing for the second. From this we infer that $f_\pm$ shows the same dependence on angles in the limit as is shown by $\psi_o \pm \bar{\psi}_o$. More generally, since

$$\psi_m \pm \psi_m = e^{iky\,\sin\theta_m}[e^{ikx\cos\theta_m} \pm e^{ikx\cos\theta_m}]$$
$$\to 2e^{iky\,\sin\theta_m}\begin{bmatrix} 1 \\ ihr\cos\theta\cos\theta_m \end{bmatrix}, \tag{65}$$

as either θ or $\theta m \to \pi/2$, we assume

$$(f_{m\nu})^- = O(\epsilon) \to 0, \quad (f_{mm})^- = O(\epsilon^2) \to 0;$$
$$(f_{m\nu})^+,\ (f_{mm})^+ \text{ nonvanishing.} \tag{66}$$

(It may also be deduced directly from theorems (49a) and (50a) that the real part of $(f_{mm})^-$ vanishes essentially to the second order of $(f_{m\nu})^-$ as $\theta_m \to \pi/2$, and also that the plus cases are finite; there is of course no question about $(f_{mm})^+$ since its real part relates to twice the scattering cross section of the cylinder.) Consequently, since $z_{m\nu} = C_m f_{m\nu}$, $C_m \to (kd\epsilon)^{-1}$ we have for all ν

$$z_{\nu m} = O(\epsilon) \to 0,\ z_{m\nu}^+ = O(\epsilon^{-1}) \to \infty;$$

$$\theta_m = \frac{\pi}{2} - \epsilon \to \frac{\pi}{2}. \tag{66a}$$

Returning to (64), we thus see that $z^- \to 0$, $z^+ \to \infty$ as $\theta \to \pi/2$, and consequently $\rho^\pm \to \pm 1$. In view of our definitions, the incident wave for the reflection grating is $\pm\psi_o$ for the two cases (corresponding to polarization $\perp$ and $\parallel$ to the generators of the bosses); hence both polarization components have undergone phase changes of π on reflection; i.e., as for a conducting dielectric plane.

The corresponding amplitudes for the transmission grating are

$$t = \tfrac{1}{2}(\rho^+ + \rho^-) = (1 - z^+z^-)/D$$
$$= [1 - C^2(g_i^2 - g_r^2)]/D,$$
$$r = \tfrac{1}{2}(\rho^+ - \rho^-) = (z^+ - z^-)/D = 2g_rC/D;$$
$$D = (1 - z^+)(1 - z^-) = (1 - g_iC)^2 - (g_rC)^2,$$

$$g_t = \bar{g}(\theta, \theta) = \bar{g}(\pi - \theta, \pi - \theta),$$
$$g_r = \bar{g}(\theta, \pi - \theta) = \bar{g}(\pi - \theta, \theta). \quad (67)$$

It may be shown that $|t|^2 + |r|^2 = 1$, and also $|t + r|^2 = 1$ in accord with the statement following (43).

If we keep two propagating modes of (62), then we obtain

$$\rho_o = (1 - z_{oo} - z_{11} - \Delta)/D, \quad \rho_1 = 2z_{o1}/\Delta;$$
$$\Delta = \begin{bmatrix} z_{oo} z_{o1} \\ z_{1o} z_{11} \end{bmatrix}, \quad D = 1 - z_{oo} - z_{11} + \Delta, \quad (68)$$

or, equivalently,

$$\rho_o = \frac{1 + x}{1 - x}, \quad \rho_1 = \frac{2z_{1o}}{(1 - z_{11})(1 - x)};$$
$$x = \frac{z_{oo} - \Delta}{1 - z_{11}} = z_{oo} + \frac{z_{o1} z_{1o}}{1 - z_{11}}, \quad (69)$$

which fulfills the energy relation $1 = |\rho_o|^2 + |\rho_1|^2 \cos\theta_1/\cos\theta_o$; this may be shown by using $z_{o1} = (\cos\theta_1/\cos\theta_o)z_{1o}$, and the fact that z is imaginary. If $\theta_1 = \pi/2 - \epsilon \to \pi/2$, then $z_{o1}^-, z_{11}^- \to 0$, so that $x^- \to z_{oo}^-$; hence $\rho_1^- \to 0$, and the result reduces to the one mode case. On the other hand, $z_{1o}^+, z_{11}^+ \to \infty$, so that

$$x^+ \to z_{oo}[1 - z_{o1}z_{1o}/z_{oo}z_{11}]$$

is imaginary; hence $|\rho_o|^2 = 1$, and $|\rho_1|^2 \cos\theta_1 = 0$. This result ρ_o^+ differs from the one mode case, in that its phase is influenced by the mode at grazing emergence. If mode 1 is nonpropagating, then x must in general be imaginary in order that $|\rho_o|^2 = 1$ (we also have the special cases $x \to 0$, ∞ and $\rho_o \to -1, +1$); hence $\bar{f}_{11}$ is imaginary, and $\bar{f}_{o1}$ is real or imaginary. From this it follows that $\bar{f}_{mm}$ is imaginary and that $\bar{f}_{m\nu}$ is real or imaginary, where m represents a surface mode; since we use this in the following only to infer that $\bar{f}_{m\nu}^2$ is real, we may consider all $\bar{f}$'s to be imaginary in order to facilitate discussion.

We now treat an arbitrarily large number of modes; we assume that the $\bar{f}$'s are at least moderately small, and imaginary, and seek series expansions of (61). If there is no mode near $\pi/2$, then $C_\nu = (kd \cos\theta_\nu)^{-1}$ is small, and we may in general expand (61) by a Neumann iteration procedure starting with $\bar{f}_{\mu o}$. Thus

$$F_\nu = \bar{f}_{\nu o} + \bar{S}C_\mu \bar{f}_{\nu\mu}F_\mu = \bar{f}_{\nu o} + \bar{S}C_\mu \bar{f}_{\nu\mu}\bar{f}_{\mu o}$$
$$+ \bar{S}\bar{S}C_\mu C_\alpha \bar{f}_{\nu\mu}\bar{f}_{\mu\alpha}\bar{f}_{\alpha o} \cdots, \quad (70)$$

which may be interpreted in terms of "orders of scattering" (if we remove the bars).

In accord with the energy theorem (37), a consistent approximation to lowest order requires that we retain $F_o \approx \bar{f}_{oo} + \bar{S}c_\mu \bar{f}_{\mu o}^2$, $F_v \approx \bar{f}_{\nu o}$; thus

$$\text{Re } F_o \approx \text{Re } \bar{S}C_\mu \bar{f}_{\mu o}^2 = -\sum C_x |\bar{f}_{\mu o}|^2,$$
$$|F_\nu|^2 \approx |\bar{f}_{\nu o}|^2; \quad (71)$$

where, in the above and in all that follows, we use the plain Σ to sum only over the propagating modes. Equivalently, we have

$$\rho_o \approx 1 + 2C_o\bar{f}_{oo} + 2C_o\bar{S}C_\mu \bar{f}_{\mu o}^2, \quad \rho_\nu \approx 2C_\nu \bar{f}_{\nu o};$$
$$|\rho_o|^2 \approx 1 + 4C_o^2|\bar{f}_{oo}|^2 + 4C_o\sum C_\mu \bar{f}_{\mu o}|^2$$
$$= 1 - 4C_o \sum_{\mu=o} C_\mu |\bar{f}_{\mu o}|^2, \quad |\rho_\nu|^2 \approx |2C_\nu \bar{f}_{\nu o}|^2. \quad (72)$$

Hence, since $(\cos\theta_\nu/\cos\theta_o)|\rho_\nu|^2 = 4C_oC_\nu|f_{\nu o}|^2$, theorem (38) is fulfilled. The analogous expressions for the transmission grating are given by

$$t_o \approx 1 + 2C_o\bar{g}_{oo} - 2C_o\sum C_\mu[|\bar{g}_{\nu o}|^2 + |\bar{g}_{\nu o}|^2],$$
$$t_\nu \approx 2C_\nu \bar{g}_{\nu o} = 2C_\nu \bar{g}(\theta_\nu, \theta_o),$$
$$r_\nu \approx 2C_\nu \bar{g}_{\nu o}' = 2C_\nu \bar{g}(\pi - \theta_\nu, \theta_o); \quad (72a)$$
$$|t_o|^2 \approx 1 + 4|C_o\bar{g}_{oo}|^2 - 4C_o\sum C_\mu[|\bar{g}_{\mu o}|^2 + |\bar{g}_{\mu o}'|^2],$$
$$|t_\nu|^2 \approx 4|C_\nu \bar{g}_{\nu o}|^2, \quad |r_\nu|^2 \approx 4|C_\nu \bar{g}_{\nu o}'|^2.$$

For simplicity we restrict the following discussion largely to ρ and $\bar{f}$, and use the terminology of the reflection grating.

We note that, even to lowest order, our results differ significantly from a single scattering approximation in that they contain $\bar{f}$ instead of f. The presence of f (unless it were imaginary) would give rise to a term $\text{Re } f_{oo}$ in (71); such a term in F_o would be incorrect physically since it represents the scattering cross section of a boss (and therefore indicates omnidirectional, or "incoherent," scattering, instead of discrete scattering losses). The numerical difference in retaining $\bar{f}$ instead of f, however, is in general small, since we have assumed a large number of propagating modes; thus using $\bar{f}$ instead of f in (71) introduces the error term $(\Sigma - \int d\mu)_p 2(\text{Re } C_\mu f_{\mu o})^2$.

If there is a propagating mode near $\pi/2$, say $\theta_m = \pi/2 - \epsilon$, then from (66) we see that $(C_m\bar{f}_{\nu m})^-$ remains finite, but that $(C_m\bar{f}_{\nu m})^+ \to O(\epsilon^{-1}) \to \infty$. Hence the above expansion suffices for F_- (provided that $\bar{f}_-$ is indeed small), but an alternative expansion is required for F_+.

We recast F_ν in a form giving convergent results for $\theta_m = \pi/2$ by suppressing mode m explicitly. We write

$$F_\nu = \bar{f}_{\nu o} + C_m \bar{f}_{\nu m}F_m + \bar{S}'C_\mu \bar{f}_{\nu\mu}F_\mu, \quad (73)$$

where the prime indicates the exclusion of mode m. Solving for F_m we obtain

$$F_m = \frac{\bar{f}_{mo} + \bar{S}'C_\mu \bar{f}_{m\mu}F_\mu}{1 - C_m \bar{f}_{mm}}, \quad (74)$$

and substituting this result back into (73) gives the reduced set of equations

$$F_\nu = p_{\nu o} + \bar{S}'C_\mu p_{\nu\mu}F_\mu,$$
$$p_{\nu\mu} \equiv \bar{f}_{\nu\mu} + \frac{C_m \bar{f}_{\nu m}\bar{f}_{m\mu}}{1 - C_m \bar{f}_{mm}}. \quad (75)$$

To lowest order in p we obtain (71) and (72) with f replaced by p.

If now $\theta_m \to \pi/2$, $C_m \to \infty$, we obtain

$$p_{\nu\mu}^+ \to \bar{f}_{\nu\mu} - \bar{f}_{\nu m}\bar{f}_{m\mu}/f_{mm}, \quad p_{\nu\mu}^- \to \bar{f}_{\nu\mu}, \quad (76)$$

where we used (66) for the behavior of $\bar{f}_{m\nu}$. Hence, in general, we may iterate

$$F_\nu^+ \to p_{\nu o} + \bar{S}C_\mu p_{\nu\mu}F_\mu^+ \quad (77)$$

essentially as for the preceding case. Thus, to lowest order

$$| p_o{}^+ |^2 \approx 1 - 4C_o \sum_{\mu=o} C_\mu | p_{o\mu} |^2,$$

$$| p_v{}^+ |^2 \approx | 2C_v p_{vo} |^2. \quad (78)$$

The ratio of the present result to that of (72),

$$\left| \frac{p_{vo}}{f_{vo}} \right|^2 = \left| 1 - \frac{\bar{f}_{vm} \bar{f}_{mo}}{\bar{f}_{vo} \bar{f}_{mm}} \right|^2, \quad (79)$$

may be greater or less than unity.

The wavelengths for which the above applies fulfill the condition

$$\sin \theta_m = 1 = \sin \theta_o + \frac{2\pi m}{h_o d},$$

$$\lambda_o = (1 - \sin \theta_o) d/m. \quad (80)$$

These, and the corresponding set for $\sin \theta_{m'} = -1$, are the Rayleigh wavelengths.[5]

Now, instead of $\theta_m = \pi/2 - \epsilon$ as above, we consider $\theta_m = \pi/2 - i\epsilon$ (i.e., the case of a surface mode near $\pi/2$); this mode, although absent in the energy relations, makes its presence felt by its modification of the propagating amplitudes. We consider first the case $F = F^+$. Ignoring temporarily the obvious "resonance" at $C_m \bar{f}_{mm} - 1 = 0$, we see from (75) that the leading term of F_v may vanish:

$$F_v \approx p_{vo} \to \bar{f}_{vo} \left(\frac{\epsilon - \epsilon_v}{\epsilon - \epsilon_m} \right);$$

$$\epsilon_v \equiv \frac{\bar{f}_{vo} \bar{f}_{mm} - \bar{f}_{vm} \bar{f}_{mo}}{ik\, d\bar{f}_{vo}}, \qquad \epsilon_m \equiv \frac{\bar{f}_{mm}}{ik\, d}. \quad (81)$$

We may write

$$I = \left| \frac{p_{vo}}{\bar{f}_{vo}} \right|^2 = \left(\frac{\epsilon - \epsilon_v}{\epsilon - \epsilon_m} \right), \quad (82)$$

where we assume $\epsilon_v < \epsilon_m$ for discussion purposes, and also that ϵ_v and ϵ_m are independent of ϵ. We find

$$\frac{\partial I}{\partial \epsilon} = \frac{2(\epsilon - \epsilon_v)(\epsilon_v - \epsilon_m)}{(\epsilon - \epsilon_m)^2}, \quad (83)$$

which vanishes at $\epsilon = \epsilon_v$ corresponding to the minimum value (81). Since

$$\sin \theta_m = \sin \left(\frac{\pi}{2} - i\epsilon \right) = \sin \theta_o + \frac{m\lambda}{d} \approx 1 + \frac{\epsilon^2}{2}, \quad (84)$$

we have

$$\frac{\partial I}{\partial \lambda} = \frac{\partial I}{\partial \epsilon} \frac{m}{d\epsilon} \to \infty; \qquad \epsilon \to 0, \qquad \lambda \to \lambda_o. \quad (83a)$$

Hence to this order of approximation, the intensity-wavelength curve is vertical at $\epsilon = 0$ (i.e., at the Rayleigh value $\lambda = \lambda_o$), and the intensity drops to zero at the distance

$$\lambda_v - \lambda_o = \frac{d\epsilon_v{}^2}{2m} = \frac{\pi \epsilon_v{}^2}{k_o(1 - \sin \theta_o)} \quad (85)$$

on the long wavelength side of λ_o. This is in accord with the "first half" of the anomaly described in the introduction.

(We note that the "two mode" formalism of (69) may also be used for the case of two propagating modes (0, and $n = -1$) plus a third nonpropagating mode

$(m = 1)$, by replacing the previous $z_{v\mu}$ by $y_{v\mu} = C_v p_{v\mu}$. Thus we see that p_n vanishes if $y_{no} = 0$, as above, and also that there is no difficulty for the case $1 - C_m \bar{f}_{mm} \to 0$, $\epsilon \to \epsilon_m$.)

We now consider the range $1 - C_m \bar{f}_{mm} \approx 0$. We substitute

$$p_{v\mu} \to \frac{C_m \bar{f}_{vm} \bar{f}_{m\mu}}{1 - C_m \bar{f}_{mm}} \quad (86)$$

into (75), and obtain

$$F_v \to f_{vm} C_m F_m, \quad (87)$$

where F_m is given in (74). Substituting this value of F_v into F_m of (74), we solve for F_m, and substitute it into (87). Thus we obtain

$$F_v \approx \frac{C_m \bar{f}_{vm} \bar{f}_{mo}}{1 - C_m \bar{f}_{mm} - C_m \bar{S}' C_\mu \bar{f}_{m\mu}{}^2} \to - \frac{\bar{f}_{vm} \bar{f}_{mo}}{\bar{S}' C_\mu \bar{f}_{m\mu}{}^2}, \quad (88)$$

which indicates a maximum in the intensity curve at the wavelength

$$\lambda_m - \lambda_o = \epsilon_m{}^2 d/2m, \qquad \epsilon_m = \bar{f}_{mm}/ikd. \quad (89)$$

The vicinity of λ_m corresponds essentially to the "second half" of an anomaly. At λ_m we have

$$\rho_o = 1 + 2C_o F_o = (\bar{S} C_\mu \bar{f}_{m\mu}{}^2 - 2C_o \bar{f}_{om}{}^2)/D,$$

$$\rho_v = 2C_v F_v = -2C_v \bar{f}_{vm} \bar{f}_{mo}/D,$$

$$D = S' C_\mu f_{\mu m}{}^2, \quad (90)$$

or, equivalently,

$$\rho_o = \frac{K - 1}{K + 1}, \qquad \rho_v = - \frac{2 \cos \theta_o \bar{f}_{vm}}{\cos \theta_v \bar{f}_{mo}(K + 1)};$$

$$K = \frac{S'}{\mu=o} \left[\frac{\cos \theta_o}{\cos \theta_\mu} \left(\frac{f_{\mu m}}{f_{mo}} \right)^2 \right], \qquad \text{Re } K = \sum_{\mu=o} [\ \ ]; \quad (91)$$

since $\sum_{v=o} | p_v |^2 \cos \theta_v / \cos \theta_o = 4\text{Re } K/| 1 + K |^2$, these results fulfill the energy principle. (If we take $\lambda_m - \lambda_o$ as a measure of the width of the anomaly, and approximate kd by $k_o d = 2m\pi/(1 - \sin \theta_o)$, then $(f_{mm}/f_{nn})^2 (n/m)^3$ indicates the relative width of anomalies corresponding to different λ_o's.)

Up to this point we considered the two "halves" of the anomaly separately; we obtain a more general form for F_v as follows. We regard $\bar{f}_{vo} + C_m \bar{f}_{vm} F_m$ of (73) as the inhomogeneous term of the equation, and iterate it once; and then proceed essentially as for (74), etc., to suppress the mth mode. Retaining only the leading term, we obtain

$$F_v \approx \bar{f}_{vo} + \frac{\bar{f}_v \bar{f}_{mo} C_m}{1 - C_m \bar{f}_{mm} - C_m \bar{S}' C_\mu \bar{f}_{m\mu}{}^2}, \qu(92)$$

which possesses the essential features of the cases of interest. (A relation of somewhat similar form was suggested by Fano[6] on the basis of a quantum mechanical analogy.)

The corresponding intensity ratio for the + case may be written

$$I_\nu = \left| \frac{F_\nu}{f_{\nu 0}} \right|^2 = \left| \frac{\epsilon - \epsilon_\nu - \Gamma}{\epsilon - \epsilon_m - \Gamma} \right|^2 ;$$

$$= \frac{(\epsilon - \epsilon_\nu - B)^2 + A^2}{(\epsilon - \epsilon_m - B)^2 + A^2} ,$$

$$\Gamma \equiv iA + B \equiv \bar{S}' \frac{(\bar{f}_{\mu m})^2}{i \cos \theta_\mu (k\,d)^2}, \quad (93)$$

where ϵ_ν and ϵ_m, as given in (81) and (82), are assumed positive. We have

$$\frac{\partial}{\partial \epsilon} I = 2 \frac{(\epsilon_\nu - \epsilon_m)[(\epsilon - \epsilon_\nu - B)(\epsilon - \epsilon_m - B) - A^2]}{[(\epsilon - \epsilon_m - B)^2 + A^2]^2} , \quad (94)$$

which vanishes when

$$\epsilon = \tfrac{1}{2}(\epsilon_\nu + \epsilon_m + 2B) \pm \tfrac{1}{2}[(\epsilon_\nu - \epsilon_m)^2 + 4A^2]^{1/2};$$

$$A = - \sum \frac{(\bar{f}_{\mu m})^2}{\cos \theta_\mu (kd)^2},$$

$$B = -S_\nu' \frac{(\bar{f}_{\mu m})^2}{\sin \gamma_\mu (kd)^2}. \quad (95)$$

If

$$\left(\frac{A}{\epsilon_\nu - \epsilon_m} \right)^2 = \left[\frac{\bar{f}_{\nu 0}}{\bar{f}_{\nu m} \bar{f}_{m 0}} \sum \frac{C_\mu}{kd} (\bar{f}_{\mu m})^2 \right]^2 \equiv M \ll 1, \quad (96)$$

and if we ignore B, then we obtain the previous cases;

$$\epsilon_{min} \approx \epsilon_\nu, \ \epsilon_{max} \approx \epsilon_m . \quad (96a)$$

(Note that ϵ_{min} depends on the spectral order ν, but that ϵ_{max} depends only on mode m.) More generally we have

$$\epsilon_{min} \approx \epsilon_\nu + B - A^2/(\epsilon_m - \epsilon_\nu),$$

$$\epsilon_{max} \approx \epsilon_m + B + A^2/(\epsilon_m - \epsilon_\nu). \quad (97)$$

The minimum intensity is $\approx |\bar{f}_{\nu 0}|^2 M$; the maximum is $\approx |f_{\nu 0}|^2/M$. We assumed $\epsilon_\nu < \epsilon_m$ for discussion purposes, so that the minimum fell at shorter wavelengths than the maximum; however, we may also have $\epsilon_\nu > \epsilon_m$, and the maximum at shorter wavelenghts. The sign of $\partial_\epsilon I$ at $\epsilon = 0$ is determined by $\epsilon_\nu - \epsilon_m$, and this indicates whether the intensity curve falls or rises to its value at λ_0.

The previous discussion applies largely for F^+ (polarization perpendicular to the elements of a reflection grating). For F^- we use (66) and replace ϵ_ν by $\epsilon^2 \beta_\nu$, etc. Thus instead of (93) we obtain

$$I_- = \left| \frac{F_\nu^-}{\bar{f}_{\nu 0}} \right|^2 = \left| \frac{\epsilon^{-1} - \beta_\nu - \Delta}{\epsilon^{-1} - \beta_m - \Delta} \right|^2 ;$$

$$\beta_\nu = \frac{\epsilon_\nu^-}{\epsilon^2}, \ \beta_m = \frac{\epsilon_m^-}{\epsilon^2}, \ \Delta = \frac{\Gamma^-}{\epsilon^2} \equiv ic + D, \quad (98)$$

where β_ν, etc., which are real and independent of ϵ, are also assumed to have the appropriate magnitudes to justify the procedure leading to (92). We replace ϵ^{-1} by β in (98) and proceed essentially as for (93):

$$\partial_\epsilon I^- \propto (\beta - \beta_\nu - D)(\beta - \beta_m - D) - c^2, \ \beta = \frac{1}{\epsilon}, \quad (99)$$

etc. It seems clear that, in general, the values of ϵ corresponding to the roots of the equation $\partial_\epsilon I^- = 0$ are not compatible with the requirement $\epsilon \ll 1$; thus only a trend towards the anomalies exists. To lowest order, $I \approx 1 + \epsilon(\beta_m - \beta_\nu)$, and consequently $F_\nu^- \to \bar{f}_{\nu 0}$, $\epsilon \to 0$; i.e., unlike the case of F^+, the original Neumann expansion (70) suffices at the Rayleigh wavelengths.

We may also treat the case where there are two modes near the plane of the grating, i.e., when $\theta_m \approx \pi/2$, $\theta_n \approx -\pi/2$; this corresponds to the so-called *double anomalies*.[4] Thus we write

$$F_\nu = \bar{f}_{\nu 0} + C_m \bar{f}_{\nu m} F_m + C_n \bar{f}_{\nu n} F_n \pm \bar{S}'' C_\mu f_{\nu \mu} F_\mu , \quad (100)$$

where the double prime indicates the exclusion of m and n, and proceed essentially as previously. Solving the pair of simultaneous equations for F_m and F_n, and then substituting the results back into (100), we obtain the reduced set of equations

$$F_\nu = T_{\nu 0} + \bar{S}'' C_\mu T_{\nu \mu} F_\mu,$$

$$T_{\nu \mu} = \bar{f}_{\nu \mu} + \bar{f}_{\nu m} L_{mn\mu} + \bar{f}_{\nu n} L_{nm\mu} ,$$

$$L_{mn\mu} = \frac{C_m[\bar{f}_{m\mu}(1 - C_n \bar{f}_{nn}) + C_n \bar{f}_{n\mu} \bar{f}_{nm}]}{(1 - C_n \bar{f}_{nn})(1 - C_m \bar{f}_{mm}) - C_m C_n \bar{f}_{mn}^2}. \quad (101)$$

(Essentially the form of the leading term $F_\nu \approx T_{\nu 0}$, but containing g instead of $\bar{f}$, was obtained originally for the transmission grating by Karp[7] through elementary considerations, and by Twersky[7] by means of an "orders of scattering series;" the C's were not the same functions we use here.)

The Rayleigh wavelengths for this case fulfill $\lambda_0 = 2d/(m - n) = -2d \sin \theta_0/(m + n)$, where n is negative, and $|n| > m$. The simplest cases correspond to normal incidence; here $C_m = C_n$, $\bar{f}_{mm} = \bar{f}_{nn}$, $\bar{f}_{n0} = \bar{f}_{m0}$, and consequently

$$T_{\nu 0} = f_{\nu 0} + \frac{C_m \bar{f}_{m0}(\bar{f}_{\nu m} + \bar{f}_{\nu n})}{1 - C_m(\bar{f}_{mm} + \bar{f}_{mn})} .$$

We may now define $\epsilon_m = (\bar{f}_{mm} + \bar{f}_{mn})/ikd$, etc., and proceed essentially as for the case of one mode near grazing. Similarly the general case offers no particular difficulties.

The treatment of this section indicates the essential features of the Wood[4] anomalies observed for perpendicular polarization. The marked effects observed by Palmer[4] for parallel polarization are not indicated by our model, and this is most likely due to our perfectly conducting "ground plane"; i.e., the plane insures that $f_{mm}^- = f(\pi/2, \pi/2)$ vanishes, and so "suppresses" the anomalies. The essential requirement for anomalies seems to be that the field of an isolated boss be nonvanishing near grazing. (We also infer that Palmer's gratings did not have a highly conducting "ground plane", from the fact that they gave large grazing modes;[4] this follows from the single scattering theorems (49a) and (50a) applied to a finite grating as a unit.)

The two polarizations compare as follows when $\theta_m \to \pi/2$: although the first Fresnel zone of the grazing mode becomes infinitely wide ($C_m \to (hd \cos \theta_m)^{-1} \to \infty$), the amplitudes of the boss approach zero ($F_m^+ \to -(\bar{f}_{m0}/c_m \bar{f}_{mm})^+ \to 0$, $F_m^- = (\bar{f}_{m0})^- \to 0$), and consequently both reflection amplitudes remain finite and nonvanishing. For the perpendicular case, the reflection amplitude becomes of the order unity ϵ $\rho_m = 2C_m F_m \to -2\bar{f}_{m0}/\bar{f}_{mm}$, and thus $|\rho_m|^2$ becomes "intenser" than all but the central mode. However, since $|\rho_m|^2 \cos \theta_m \to 0$, this mode derives no time-averaged power from the incident wave in the limit $\theta_m = \pi/2$. (Note that the

value of $|\rho_m|$ corresponding to $C_m f_{mm} = 1$; i.e., $\rho_m = -2 f_{mo}/S C_\mu (f_{\mu m})^2$, is smaller than the present value.) Although the corresponding functions for the two polarizations are of the same order in ϵ, the perpendicular values are in general much larger. Thus if we assume that $f^-/\epsilon = O(\overline{f^+}) < 1$, then

$$| \rho^+/\rho^- |^2 \rightarrow (kd/f^+)^2 >>> 0.$$

For simplicity, we restricted discussion of the anomalies to the case of the reflection grating. However, the analogous bands may also occur for the transmission grating. Thus corresponding to (93) and (98), we now have

$$\left| G\left(\begin{matrix} \theta_\nu \\ \pi - \theta_\nu \end{matrix}, \theta_o \right) \right|^2$$
$$\simeq \tfrac{1}{4} \left| f_{\nu o}{}^+ + \frac{\epsilon - \epsilon_\nu - B - iA}{\epsilon - \epsilon_m - B - iA} \pm f_{\nu o}{}^- \right|^2 \quad (103)$$

Because of the "background" provided by F^-, the bands will not be as well defined as in (93); however, the pronounced maxima of F^+ should persist.

We simplified matters by restricting discussion to symmetrical nonabsorbing scatterers, but the case where f is complex offers no particular difficulties. The material may be applied to a specific grating provided that f is known explicitly. In particular, we apply it to the circular cylinder in the following.

VI. Grating of Circular Cylinders

In general, the scattering amplitude of a circular cylinder may be written as

$$g(\theta, \theta_o) = \sum_{n=-\infty}^{\infty} A_n e^{in(\theta - \theta_o)} = \sum A_n \cos n(\theta - \theta_o), \quad (104)$$

where, e.g.,

$$A_n{}^- = -\frac{J_n(ka)}{H_n(ka)}, \quad \overline{\Psi} = 0 \text{ at } r = a;$$
$$A_n{}^+ = -\frac{J_n'(ha)}{H_n'(ha)}, \quad \partial_r \Psi = 0 \text{ at } r = a; \quad (105)$$

these correspond respectively to $\Psi = E_z$, H_z for the electromagnetic case of a perfect conductor (or to a free, rigid surface in acoustics); here the prime indicates differentiation with respect to argument. We note that the series of (104) converges for complex as well as for real values of the angles.

We substitute (104) and a similar series for g into (55). Integrating and collecting the coefficients of $e^{in\theta}$, we obtain the transformed scattering amplitude

$$g(\theta, \theta_o) = \sum_{n=-\infty}^{\infty} \overline{A}_n e^{in(\theta - \theta_o)}$$
$$= \sum A_n \cos n(\theta - \theta_o), \quad A_n = \frac{A_n}{1 + A_n}. \quad (106)$$

From (104) and (49a) we obtain $-\operatorname{Re} A_n = |A_n|^2$ for the general case of nonabsorption; for perfect conductors this result follows directly from the forms of (105). Consequently, for nonabsorbers

$$g(\theta, \theta_o) = i \sum_{n=-\infty}^{\infty} \frac{\operatorname{Im} A_n}{|1 + A_n|^2} \cos n(\theta - \theta_o) \quad (106a)$$

is imaginary [as shown for the more general case in (57a)].

In particular, for the perfect conductor, we use (105) and obtain

$$A = i \frac{J_n(ka)}{N_n(ka)}, \quad A^+ = i \frac{J_n'(ka)}{N_n'(ka)} \quad (107)$$

We now consider perfect conductors for $ka \ll 1$. To lowest order in ha we obtain for E polarization parallel, perpendicular to the axis

$$g_{||}(\theta, \theta_o) = -\frac{i\pi}{2 \log (2/\gamma ka)} + O(ka)^2, \quad \gamma = 1.781 \cdots ;$$

$$g_\perp(\theta, \theta_o) = -\frac{i\pi(ka)}{4}{}^2 [1 - 2 \cos (\theta - \theta_o)]; \quad (108)$$

to this order of approximation, $g = g$.

We first use these expressions to determine the reflection and transmission coefficients of a grating such that $ka << kd < 1$. For simplicity, we apply the heuristic form (67) (which neglects the surface waves) directly to $g_\perp$; on the other hand, we first modify $g_{||}$ to take into account the major effect of the surface wave. Thus we replace $g_{||}$ by

$$\frac{g_{||}}{1 + ig_{||}I} \approx -\frac{i\pi}{2 \log (d/a2\pi)} \equiv g_i, \quad (109)$$

where we approximated I of (127) of the Appendix by its logarithmic term.

Thus substituting g_i and $g_\perp$ into (67), we approximate the transmission and reflection coefficients for a grating of perfectly conducting wires such that $ka << kd < 1$ by,

$$T_{||}(\theta) = |t_{11}|^2 \approx \left| \frac{1}{1 + iq} \right|^2 = \frac{1}{1 + q^2},$$
$$R_{||}(\theta) = |r_{11}|^2 \approx \left| \frac{-iq}{1 + iq} \right|^2 = \frac{q^2}{1 + q^2};$$
$$q = \frac{\lambda}{2d \cos \theta \log \left(\frac{d}{2a\pi} \right)}.$$

$$T_\perp(\theta) = |t_\perp|^2 \approx \left| \frac{1 - (Q/2)^2[(1 + 2 \cos 2\theta)^2 - 1]}{1 - iQ + (Q/2)^2[(1 + 2 \cos 2\theta)^2 - 1]} \right|^2$$
$$\approx \frac{1 - Q^2[(1 + 2 \cos 2\theta)^2 - 1]}{1 + Q^2},$$

$$R_\perp(\theta) = |r_\perp|^2 \approx \left| \frac{-iQ(1 + 2 \cos 2\theta)}{1 - iQ + (Q/2)^2[(1 + 2 \cos 2\theta)^2 - 1]} \right|^2$$
$$\approx \frac{Q^2(1 + 2 \cos 2\theta)^2}{1 + Q^2}; \quad Q = \frac{\pi^2 a^2}{\lambda d \cos \theta}. \quad (110)$$

Note that not only is $|t|^2 + |r|^2 = 1$, but also $|t + r| = 1$; thus both energy theorems are fulfilled. (The parallel case has been treated numerically without the restriction $a << d$ by Honerjäger[7] and others.)

The case of normal incidence was originally treated by Lamb,[3] and Gans,[3] by means of a "static approximation". Our results for the $||$ case reduce to Gans' at this special angle; however, our results for the $\perp$ case for $\theta = 0$ are substantially different. Thus we have

$$T_\perp(0) = \frac{1 - 8Q^2}{1 + Q^2}, \quad R_\perp(0) = \frac{9Q^2}{1 + Q^2}; \quad Q = \frac{\pi^2 a^2}{\lambda d}, \quad (111)$$

and they have

$$T = \frac{1}{1 + 4Q^2}, \; R = \frac{4Q^2}{1 + 4Q^2} . \quad (112)$$

Hence our value for K/T is more than twice theirs (essentially 9/4), and this factor should be measureable. We believe that the discrepancy arises from their neglect of the monopole term of the single cylinder; i.e., from Lamb's results it follows that the "scattering amplitude" of the grating (proportional to our $G(\theta, \theta_o)$) differs only in sign for the forward- and back-scattered directions. More explicitly, Lamb's treatment is based on the flow of a fluid past a rigid cylinder; for this case the scattering amplitude (which corresponds to that of a simple diopole or double source; see Lamb,[3] p. 76) is proportional to $2 \cos \theta$. On the other hand, we see from (108), that when a wave excites a cylinder, the lowest-order terms include both a monopole and a dipole and that the amplitude is proportional to $1 + 2 \cos \theta$. Thus for the wave problem, the scattered intensity is proportional to 9 and 1 in the back and forward directions; for the fluid flow problem, the intensity is proportional to 4 in both directions; essentially this difference is indicated in the above two approximations of the multiple scattering problem.

In general we substitute $g_{\parallel}$ into (56) and obtain

$$G_{\parallel} = g_{\parallel}(1 + G_{\parallel}SC_\nu] = \frac{g_{\parallel}}{1 - 2g_{\parallel}SC_\nu} ,$$

$$C_\nu = \frac{1}{kd \cos \theta_\nu} , \quad (113)$$

where SC_ν is discussed in the Appendix. The Wood anomalies are particularly simple for this case. Thus if there is one mode (m) near grazing (propagating or nonpropagating), then

$$| t_\nu |^2 = | r_\nu |^2 = | 2C_\nu G |^2 \rightarrow \frac{4\epsilon^2}{\cos^2 \theta_\nu} \rightarrow 0, \; t_\nu \neq t_o ;$$

$$| t_o |^2 = | 1 + 2C_oG |^2 = 1 + [(2C_o)^2 - 8C_o \sum C_\nu] | G |^2$$

$$\rightarrow 1 + \left(\frac{4}{\cos^2 \theta_o} - \sum \frac{8}{\cos \theta_o \cos \theta_\nu} \right) \epsilon^2 \rightarrow 1, \quad (114)$$

where the sum includes only propagating terms. Here $\epsilon^2 = 2 m | \lambda - \lambda_o | / d$, and $\lambda_o = (1 - \sin \theta_o) d/m$ is the Rayleigh wavelength. If there are two grazing modes, m and $n = - | n |$, then we replace ϵ by $\epsilon\epsilon'/(\epsilon + \epsilon')$ where $\epsilon'^2 = \epsilon^2 | n | /m$; for this case $\lambda_o = 2d/(| n | + m)$. The Wood anomalies are thus simple minima, the energy "missing" in the spectral orders "appearing" in the forward transmitted wave; i.e., the grating is transparent for λ_o.

For perpendicular polarization, we write

$$g_\perp(\theta, \theta_o) = \tfrac{1}{2}(f_+ + f_-),$$

$$f_\pm = - \frac{i\pi(ka)^2}{2} \begin{bmatrix} 1 - 2 \sin \theta \sin \theta_o \\ - 2 \cos \theta \cos \theta_o \end{bmatrix}, \quad (115)$$

where $f_\pm$ are the components of g symmetric and antisymmetric with respect to reflection of one angle. Before discussing this case, we introduce the analogous functions for the reflection grating of perfectly conducting semicylinders on a perfectly conducting plane.

The transformed amplitudes for the reflection problem are obtained by superposition of appropriate forms of (106). Thus

$$\bar{f}_\pm(\varphi, \pi - \theta) = \sum \bar{A}_n^\pm e^{in\varphi}[e^{-in\theta} \pm e^{in(\pi - \theta)}]$$

$$= 2 \sum \bar{A}_n^\pm \frac{\cos}{\sin} \left[m \left(\frac{\pi}{2} - \varphi \right) \right] \frac{\cos}{\sin} \left[m \left(\frac{\pi}{2} - \theta \right) \right]. \quad (116)$$

(Note that here $\bar{f}_+$ is the symmetric component of $g_\perp$, and that $\bar{f}_-$ is the antisymmetric component of $g_{\parallel}$; i.e., the two present f's satisfy different boundary conditions on the cylinders as well as on the planes.) The limiting behavior of (116) for $\varphi, \theta \rightarrow \pi/2$ is in accord with (66).

If $ka << 1$, then to the order of $(ka)^2$ we obtain (115) with the sign of f_- reversed; we refer to the function for the boss as $f_-' = -f_-$ and treat both cases simultaneously.

For perpendicular polarization, f_+, the discussion of (93) applies. For the present case we have for $\theta_m = \pi/2 - i\epsilon$,

$$\epsilon_m = \frac{\pi(ka)}{2kd}^2 \equiv Q << 1,$$

$$\epsilon_\nu = \frac{2Q(1 - \sin \theta_\nu)(1 - \sin \theta_o)}{1 - 2 \sin \theta_\nu \sin \theta_o} ,$$

$$\Gamma = iQ^2\bar{S}' \frac{(1 - 2 \sin \theta_\mu)^2}{\cos \theta_\mu} , \quad (117)$$

where we ignored terms of the order ϵ^2Q in ϵ_m, etc. Using $\sin \theta_\nu = \sin \theta_o + 2\nu\pi/kd$, and the above, we may compute $I = | F_\nu/f_{\nu o} |^2$ of (93) as a function of λ for a given spectral order. We note that the form of the anomaly for a given wavelength range may be quite different in its different spectral orders.

For simplicity, we neglect the remaining surface terms in the following. The intensity ratio at the Rayleigh wavelength is given by

$$I \approx \left(\frac{\epsilon_\nu}{\epsilon_m} \right)^2 = 4 \left| \frac{(1 - \sin \theta_\nu)(1 - \sin \theta_o)}{1 - 2 \sin \theta_\nu \sin \theta_o} \right|^2 . \quad (118)$$

The minimum and maximum values are approximately

$$I_{min} \approx I_{max}^{-1} \approx \left| \frac{\text{Im } \Gamma}{\epsilon_m - \epsilon_\nu} \right|^2$$

$$= Q^2 \left[\frac{(1 - 2 \sin \theta_\nu \sin \theta_o)}{(1 - 2 \sin \theta_\nu)(1 - 2 \sin \theta_o)} \right.$$

$$\left. \sum \frac{(1 - 2 \sin \theta_\mu)^2}{\cos \theta_\mu} \right]^2 \quad (119)$$

and they occur approximately at $\epsilon = \epsilon_\nu$, ϵ_m respectively. (Note that since $\epsilon = | \epsilon |$, only positive values of ϵ_ν of (118) give minima.) Using (89) and (118) we find that the maximum is displaced from the Rayleigh wavelength by

$$\lambda_m - \lambda_o \sim \frac{Q^2d}{2m} \sim \frac{m(\pi a)^4}{2d^3(1 - \sin \theta_o)^2} , \quad (120)$$

where we replaced k^2 by k_o^2.

For parallel polarization, f_-', the discussion of (98) applies. Here $\beta_\nu = \epsilon_\nu^-/\epsilon^2$ is identically zero, and

$$\beta_m = \frac{\epsilon_m^-}{\epsilon^2} = 2Q, \; \Delta = \frac{\Gamma^-}{\epsilon^2} = -iQ^2S' \cos \theta_\mu . \quad (121)$$

We have

$$I = \left| \frac{1 - \epsilon\Delta}{1 - (2Q + \Delta)\epsilon} \right|^2. \qquad (122)$$

The maximum of this expression occurs for

$$\epsilon \approx (2Q)^{-1} >> 1,$$

but this is not compatible with the original conditions. To the present order of approximation we have

$$I \approx 1 + 4Q\epsilon, \ \epsilon \to 0 \ ; \qquad (123)$$

thus there are no anomalies for this component.
The above is for f_-'; i.e., for the reflection grating; for $f_- = -f_-'$ of the transmission grating, we change the sign of β_m.

The transmission grating may be treated by using the above values and the form of (103):

$$\left| G\left(\frac{\theta_\nu}{\pi - \theta_\nu}, \theta_o \right) \right|^2$$

$$\approx (Q^2/4) \left| (-1 + 2 \sin \theta_\nu \sin \theta_o) \frac{\epsilon - \epsilon_\nu - \Gamma}{\epsilon - \epsilon_m - \Gamma} \right.$$

$$\left. + 2 \cos \theta_\nu \cos \theta_o \right|^2. \qquad (124)$$

Consequently much of the detail of the anomaly in the symmetric component will be lost through interference, but the maxima should persist.

VII. Acknowledgment

We are grateful to Professor F. Reiche for suggesting the problem, and to Professor J. B. Keller, and Professor M. Schiffer for many fruitful discussions.

Appendix: Isotropic Scatterers

The special case of isotropic scatterers (e.g., $\Phi_o = 0$, $ka \to 0$, $g \to -J_o(ka)/H_o(ka)$, for a circular cylinder) was treated in detail by Honerjäger.[7] For this case g and G are independent of angles, and consequently (32) reduces to

$$G = g[1 + 2GSC_\nu] = \frac{g}{1 - 2gSC_\nu}, \qquad (125)$$

where both $\sum C_\nu$ and $\int C_\nu \, d\nu$ of SC_ν diverge individually. The real part of $2 \int C_\nu \, d\nu$ equals unity [see (53)]; in the imaginary part we put $w = \nu + kd \sin \theta_o/2\pi$, $w \to \infty$, and write

$$-\frac{2i}{\pi} \int_{kd/2\pi}^{w} \frac{dw}{[w^2 - (kd/2\pi)^2]^{1/2}} = -\frac{2i}{\pi} \log \frac{4\pi w}{kd}$$

$$= \frac{2i}{\pi} \left[\log \frac{\gamma kd}{4\pi} - \sum_{\nu=1}^{w} \frac{1}{\nu} \right], \ w \to \infty, \quad (126)$$

where $\gamma = 1.781 \cdots$; here we used $\log w = \sum_1^w \frac{1}{\nu} - \log \gamma$ (as in *MO*, p 2). Hence writing $\nu\pm$ for the limiting propagating modes [see discussion after (21)] we obtain

$$2SC_\nu = 2 \left(\sum_{-\infty}^{\infty} - \int_{-\infty}^{\infty} d\nu \right) C_\nu$$

$$= \sum_{\nu-}^{\nu+} 2[(kd)^2 - (kd \sin \theta_o + 2\nu\pi)^2]^{-1/2} - 1 - iI ; \quad (127)$$

$$I = 2(I_1 + I_{2-} + I_{2+}),$$

$$I_1 = \frac{1}{\pi} \left[\log \frac{\gamma kd}{4\pi} - \frac{1}{2} \left(\sum_1^{\nu+} + \sum_1^{|\nu-|} \right) \frac{1}{\nu} \right],$$

$$I_{2\pm} = \sum_{|\nu\pm|+1}^{\infty} \left\{ [(2\nu\pi \pm kd \sin \theta_o)^2 - (kd)^2]^{-1/2} - (2\nu\pi)^{-1} \right\},$$

where the final sum is essentially of the form $\Sigma \, (\nu)^{-3}$, and therefore converges rapidly. If $| \nu\pm | >> 1$, then I may be neglected; on the other hand if $kd \to 0$, then I may be approximated by the logarithmic term.

An alternative form of (127) follows from the physical statement of the problem; i.e., from the fact that an isotropic scatterer is an elementary line source whose radiation characteristics are completely specified by H_o. Thus the first equality of (125) may be interpreted as stating that the amplitude G equals a term arising "directly" from the incident wave, plus a term arising from the waves originating at the remaining scatterers. Both terms are proportional to g, the response of an isolated scatterer. The second is also proportional to the multiple scattered amplitude G, as is required for self-consistency, and to a factor which must be equal to the distribution of the remaining elementary sources. Thus

$$2SC_\nu = 2 \sum_{s=1}^{\infty} H_o(skd) \cos (skd \sin \theta_o). \qquad (128)$$

The equality of the right sides of (128) and (127) is well known (MO, pp 59, 60); the usual derivation of the more convergent representation (127) is based on regarding (128) as the appropriate limit of Σ of (20) minus H_o, and this was essentially implicit in our procedure.

We note that the real part of (128) may be deduced directly from the energy theorems (36) and (36a). For isotropic scatterers these theorems state

$$\text{Re } G + 1 \,|\, G \,|^2 \sum_{\nu-}^{\nu+} C_\nu = 0, \ \text{Re } g = -|\, g \,|^2. \qquad (129)$$

These relations, together with G as given in (125), give directly

$$2 \text{ Re } SC_\nu = 2 \sum_{\nu-}^{\nu+} C_\nu - 1.$$

The generalization of (128) for arbitrary circular cylinders, follows from an initial expansion of all waves in terms of complete sets of cylindrical functions. Using Reiche's result,[7] we may express the total field in the same form as (23) and (24); for this case the multiple scattered amplitude is given by [13]

$$G(\theta, \theta_o) = \sum_{n=-\infty}^{\infty} B_n(\theta_o) e^{in\theta} \qquad (130)$$

$$B_n = A_n \left\{ e^{-in\theta_0} \right.$$

$$+ \sum_{m=-\infty}^{\infty} B_m \sum_{s=1}^{\infty} H_{n-m}(skd)[e^{iskd \sin \theta_0}(-1)^{n-m}$$

$$\left. + e^{iskd \sin \theta_0}] \right\},$$

where A_n and B_n are the "single" and "multiple scattered coefficients"; e.g., $A_n = -J_n(ka)/H_n(ka)$ for $\Psi = 0$ at $r = a$. The case of normal incidence was treated by Ignatowsky.[7]

Measurement and Analysis of Instantaneous Radio Height-Gain Curves at 8.6 Millimeters over Rough Surfaces

A. W. STRAITON† AND C. W. TOLBERT†

Summary—By the use of an array of ten vertically-spaced antennas and a rotating wave guide switch, a portion of the height-gain pattern for a short radio path was obtained as a function of time for a wave length of 8.6 millimeters.

In the analysis of the data taken across a small lake, the reflection from the water is assumed to be made up of two components. One component is a constant value equal to the median signal received at the antennas over the sampling period and the other component is a variable signal of the proper phase and magnitude to give the measured total signal at each instant.

The angle of arrival, phase and magnitude of the fluctuating signal are obtained for a short sample of data and their characteristics described.

INTRODUCTION

IN STUDIES of the reflection characteristics of 8.6 millimeter radio waves,[1-3] it was found that a direct wave and a single constant ground reflected wave were not adequate to explain the nature of the resulting height-gain curves. In a previous paper[4] the resulting signal for propagation over a rough water surface was studied in terms of a reflected wave whose magnitude and phase were variable.

Although the major features of observed signals on two radio paths were explained by this method, it was realized that this approach did not provide a complete picture of the reflection from a variable rough surface.

In order to shed more light on this problem, an antenna system was constructed for obtaining a portion of a height-gain curve in a very short interval of time. This report describes this antenna system and the method of analysis of the first samples of data obtained with it.

ANTENNA SYSTEM AND RECEIVING EQUIPMENT

The chief features of the system were a ten antenna array, a rotary waveguide switch, and a heterodyne receiving system. The antennas as used for this experiment were circular horns with a length of $5\frac{1}{2}$ inches and a mouth opening of $1\frac{1}{8}$ inches. The antennas were fixed rigidly in a frame with a separation between antennas of 4 inches and are shown in Fig. 1. The antenna beam-
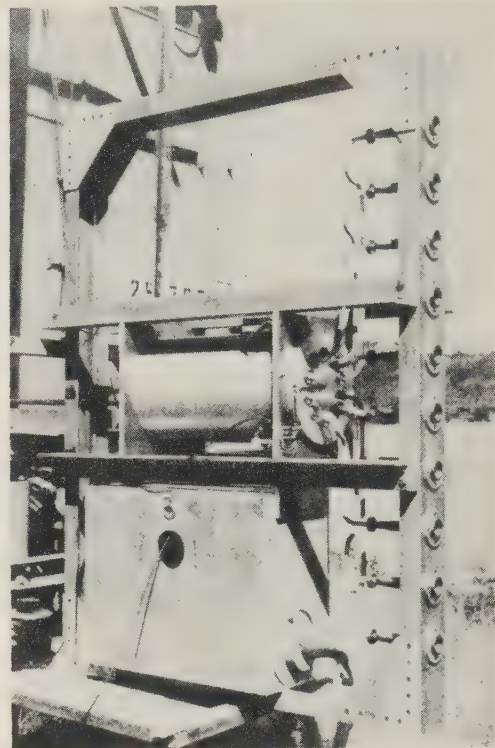
Fig. 1—View of antenna system, switch, and receiver.
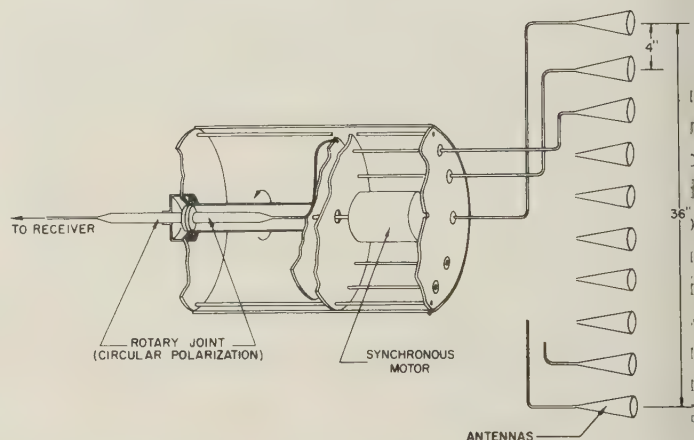
Fig. 2—Schematic drawing of waveguide switch.

width was 10 degrees and horizontal polarization was used throughout the test.

The rotary waveguide switch is housed in the cylinder directly back of the antennas in Fig. 1. This switch is shown schematically in Fig. 2. The rotary arm turned at twelve revolutions per second, thus sampling each antenna twelve times per second.

† Elec. Engrg. Res. Lab., Univ. of Texas, Austin, Tex.
[1] A. W. Straiton, A. H. LaGrone, C. W. Tolbert, C. E. Williams, "Preliminary Study of the Reflection of Millimeter Radio Waves from Fairly Smooth Ground," Elec. Engrg. Res. Lab., University of Texas, Rep. No. 60; February 29, 1952.
[2] A. W. Straiton, J. R. Gerhardt, A. H. LaGrone, C. W. Tolbert, "Reflection of Centimeter and Millimeter Radio Waves from the Surface of a Small Lake," Elec. Engrg. Res. Lab., University of Texas, Rep. No. 63; May 15, 1952.
[3] A. H. LaGrone, and C. W. Tolbert, "Reflection Studies of Millimeter and Centimeter Radio Waves for Gulf of Mexico Paths," Elec. Engrg. Res. Lab., University of Texas, Rep. No. 64; October 31, 1952.
[4] A. H. LaGrone, A. W. Straiton, H. W. Smith, "Synthesis of Radio Signals on Overwater Paths," Elec. Engrg. Res. Lab., University of Texas, Rep. No. 71; April 30, 1954.
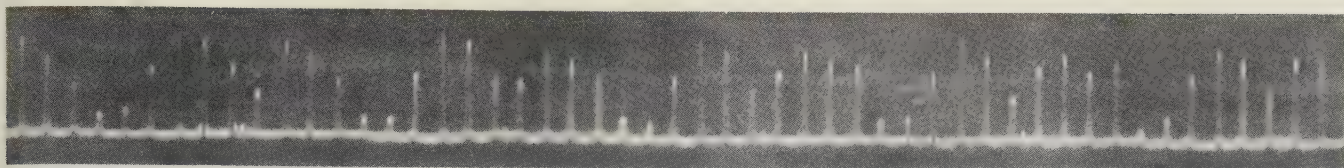
Fig. 3—Sample of original recording.

The receiver was of the heterodyne type using a QK291 klystron as a local oscillator. The output of this receiver was applied to the vertical plates of an oscilloscope with a photographic attachment. During recording intervals, the photographic film was drawn continuously through the camera and the pulses resulting from switching succeeding antennas to the input of the receiver were recorded consecutively on the film. The envelope of these pulses provided the height-gain curves for each interval of one-twelfth of a second. A sample of the original recording for several cycles is shown in Fig. 3. A marker signal indicated the transition between the top and bottom antennas. The antenna system was adjusted so that equal signals were received at each antenna with the array turned horizontally. Relative calibrations were made by attenuating the transmitted signal in known steps and comparing the levels at the various antennas. Minor corrections thus obtained were later applied to the recorded data.

The transmitter employed another QK291 klystron as a cw source with an output of approximately 20 milliwatts.

TRANSCRIPTION OF DATA

For transcription purposes, the data were projected on a screen on which there was mounted a movable cursor. As this cursor was moved across the screen to follow the data, it drove a potentiometer which controlled the voltage applied to an Esterline-Angus graphic recorder. By following the signal indicated for a single antenna, a plot of signal strength versus time for the given antenna was reproduced on the E-A recording paper. This paper was then passed through a totalizer and the statistical distribution of the signal for each antenna was obtained. This distribution was plotted on Gaussian probability paper and the points were found to fall approximately along a straight line. The median values of the individual signals were read from this line.

The data that can be obtained from this transcription are the median signal for a given run for each antenna and the magnitude of the signal at each antenna in each frame. These data will be used later in the analysis.

DESCRIPTION OF MEASUREMENT CONDITIONS

The data analyzed later in this paper were taken on a radio path over Lake Austin in November, 1954.
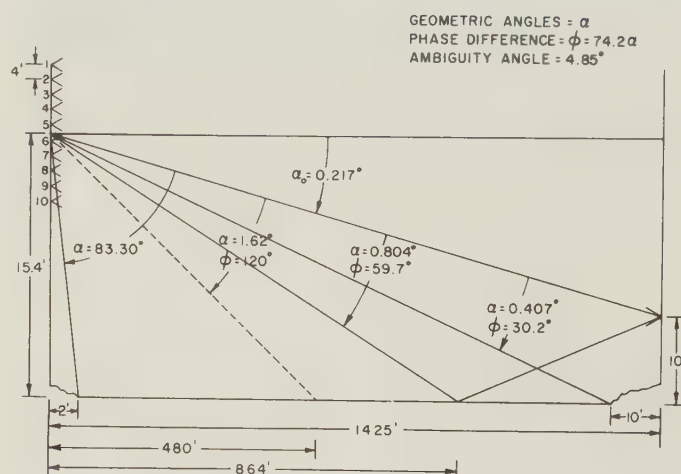
Fig. 4—Path geometry for run analyzed.

The transmitter was operated at heights of 5 feet and 10 feet above the water surface and approximately 10 feet back from the edge of the water. The receiver on the opposite shore of the lake was 1,425 feet from the transmitter and approximately 2 feet back from the water's edge. The center of the antenna array was operated at heights of 5 feet and 15.4 feet above the water surface and the antennas were spaced 4 inches apart. For reference purposes, the antennas are numbered consecutively from the top. The angles to the transmitter, to the geometric reflecting point and to the two edges of the lake are shown in Fig. 4 for the arrangement used to take the data analyzed in this report. These angles are measured from the center of the array, but they would vary only slightly from antenna to antenna.

A number of measurements were made on this path with mean angles of separation between the direct and reflected ray of 0.5 degrees and 1.0 degrees for various water surface roughness. The analysis will be primarily concerned, however, with a 100 second run taken under rough water conditions using the geometry shown in Fig. 4. The wind speed at the time was approximately 20 miles per hour producing small, choppy waves with an approximate height of 4 inches from maximum to minimum. Height markers on a pole were photographed but the position of pole near shore did not appear to be representative of the wave surface. The water surface condition, similar to that at the time of the measurement, is seen in the background of Fig. 5.

Fig. 5—View of water surface.

Vector Model of Signal Strength

In considering the characteristics of the signal received at a given antenna, let us assume that the total signal, T, is made up of a direct wave, D, a constant reflected wave, C, and a variable reflected wave, I. The combination of D and C remains constant at each antenna during the run, but varies from antenna to antenna. The components are shown in two sequences of the vectors in Figs. 6(a) and 6(b). The variable com-

(a)

(b)
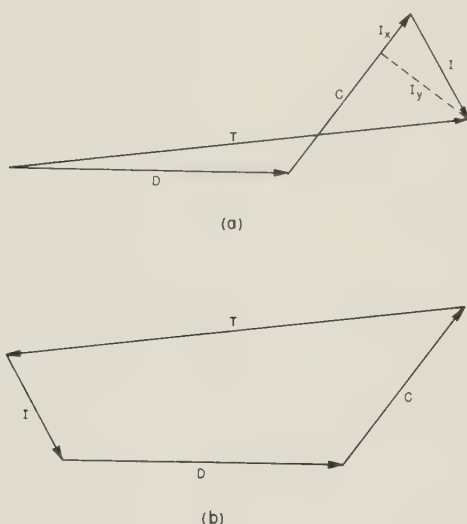
Fig. 6—Vector model.

ponent of the signal may be broken into two components, one, I_x, in-phase and the other, I_y, 90 degrees out-of-phase with the constant reflected wave. If the amplitude of the constant reflected component is in error, this would be compensated for by a shift in the value of I_x.

Solution for Variable Component

In the solution for the variable component of signal strength, the data available are the instantaneous and the median values at the ten antennas and the geometry of the path. The basic assumptions in the analysis are as follows:

1. That the height-gain function obtained from the median signals at the ten antennas can be represented by combining the direct wave and a single plane reflected wave in the proper magnitude and phase.

2. That the instantaneous variable signal can be approximated by a plane wave over the interval covered by three antennas. This plane wave may be described at a given instant by its magnitude, its angle of arrival, and its phase relative to the direct or reflected wave.
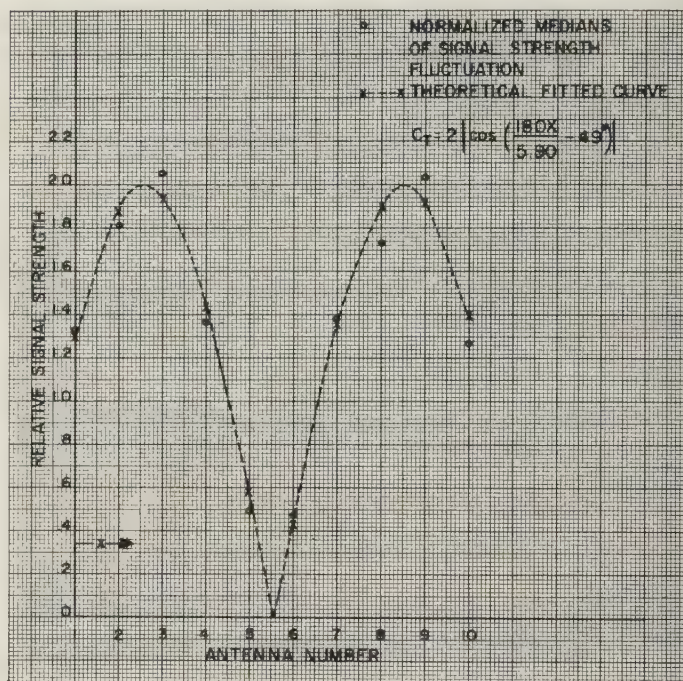
Fig. 7—Height gain curve for median data.

To the median values of the signal levels at the ten antennas, a height-gain curve was fitted as shown in Fig. 7. It is noted that a very good fit is obtained by using a reflection coefficient of one and a phase separation between the reflected waves received by two adjacent antennas of 60.0 degrees; i.e., a complete lobe includes 360/61.0 or 5.90 antenna spacings. This change in phase between the reflected wave at adjacent antennas is in agreement with that calculated from the path geometry. It is noted that the two waves assumed provide a good fit of the measured data except for antenna 8. This discrepancy is believed to be due to calibration errors for this antenna. Only the instantaneous signals received at antennas 4, 5, 6, and 7 are used in the analysis for the variable component after the height-gain has been obtained from the median values.

If the direct and the constant reflected waves are drawn in their proper phase relationship as shown in Fig. 6(b), the variable vector for a given antenna and a given frame will have one end at the origin and one end at the arc of a circle with radius T drawn from the end of the C vector.

If now, the direct wave vector for the three adjacent antennas, D_5, D_6, and D_7, are superimposed as shown in Fig. 8, the three constant reflected components, C_5, C_6, and C_7, will be as shown. The total signals for the three antennas, T_5, T_6, and T_7, for a given frame are used as radii and arcs of circles are drawn with center at the end of C_5, C_6, and C_7, respectively, as shown in Fig. 8.
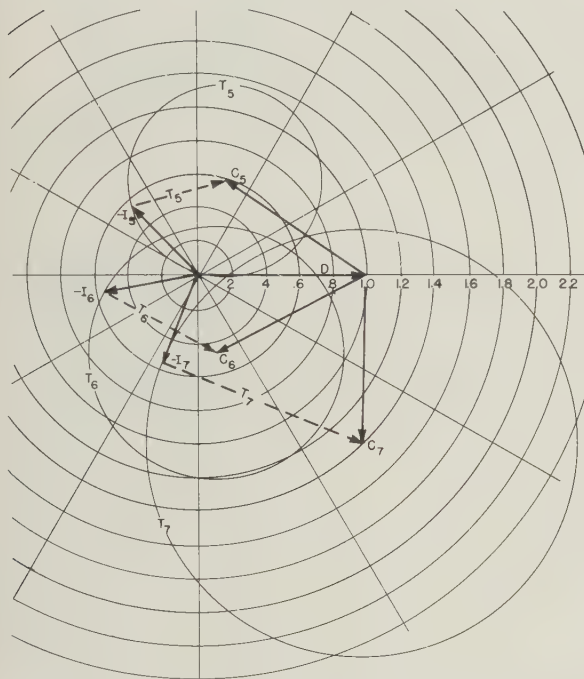
Fig. 8—Vector diagram for solution for variable component.

Under the plane wave assumption, the variable components of reflected wave, I_5, I_6, and I_7, will be equal in magnitude and uniformly spaced. The solution for I_5, I_6, and I_7 is obtained by selecting one of the concentric circles on which equal arcs are cut off by the circles formed with T_5, T_6, and T_7 as radii. From the chart, the common magnitude of I_5, I_6, and I_7, their phases relative to each other and their phases relative to the direct or constant reflected waves may be measured.

Results of Analysis

Seventy-five consecutive frames were analyzed out of a run of approximately 1,200 frames or 100 seconds. This set of seventy-five frames represented a time interval of about six seconds.
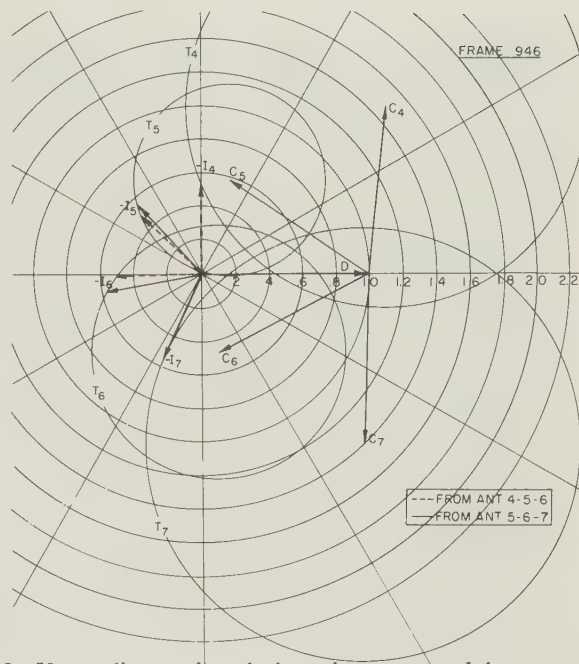
Fig. 9—Vector diagram for solution using two sets of three antennas.

In order to provide a check on the plane wave assumption, the analysis was carried through separately using antennas 4, 5, and 6 and using antennas 5, 6, and 7. The nomogram for the solution of one frame using the two sets of antennas is shown in Fig. 9. Data for all the cases obtained from this analysis are summarized in Figs. 10, 11, 12, and 13. Of the 75 frames considered, good agreement between analysis for the two sets of antennas was found in 28 cases. Fairly good agreement was obtained in 25 cases. A solution was obtained for only one set of antennas in 12 cases and a solution was obtained with neither set in 11 cases. The magnitude of I is plotted in Fig. 10 as a function of time with the vertical line joining the solutions obtained with the two sets of antennas. It is noted that the cases where no satisfactory solutions were obtained all fall in the part of the cycle when the magnitude of I is near a minimum. This would indicate that all of the undefined cases are those for which the variable component is very small. In the statistical analysis given later, the magnitude for these cases is taken as zero.

The phase difference between adjacent antennas may be converted to angle of arrival by considering the wave length and the separation of the antennas. The conversion factor by which the phase difference is to be divided to obtain angle of arrival is 74.2. The corresponding ambiguity angle is 4.85 degrees. Since this angle is fairly large compared to the geometric angles involved, it is felt that this ambiguity would cause no confusion.

The angle of arrival is plotted in Fig. 11 as a function of time with the two solutions again joined by vertical lines.
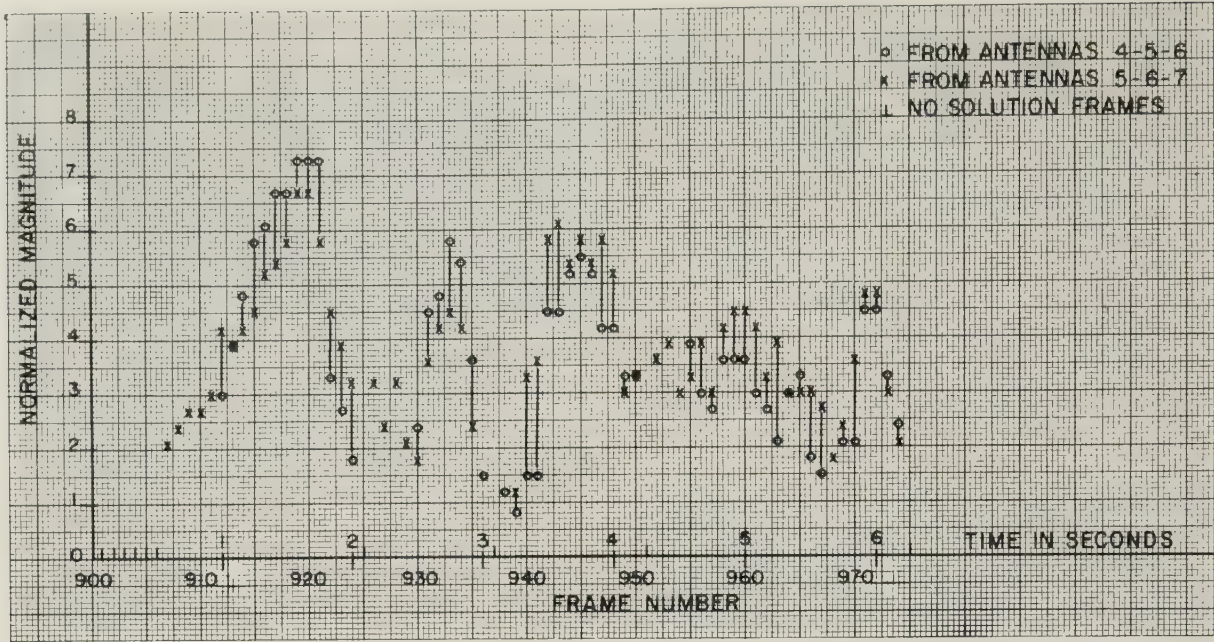
Fig. 10—Magnitude of I as a function of time.
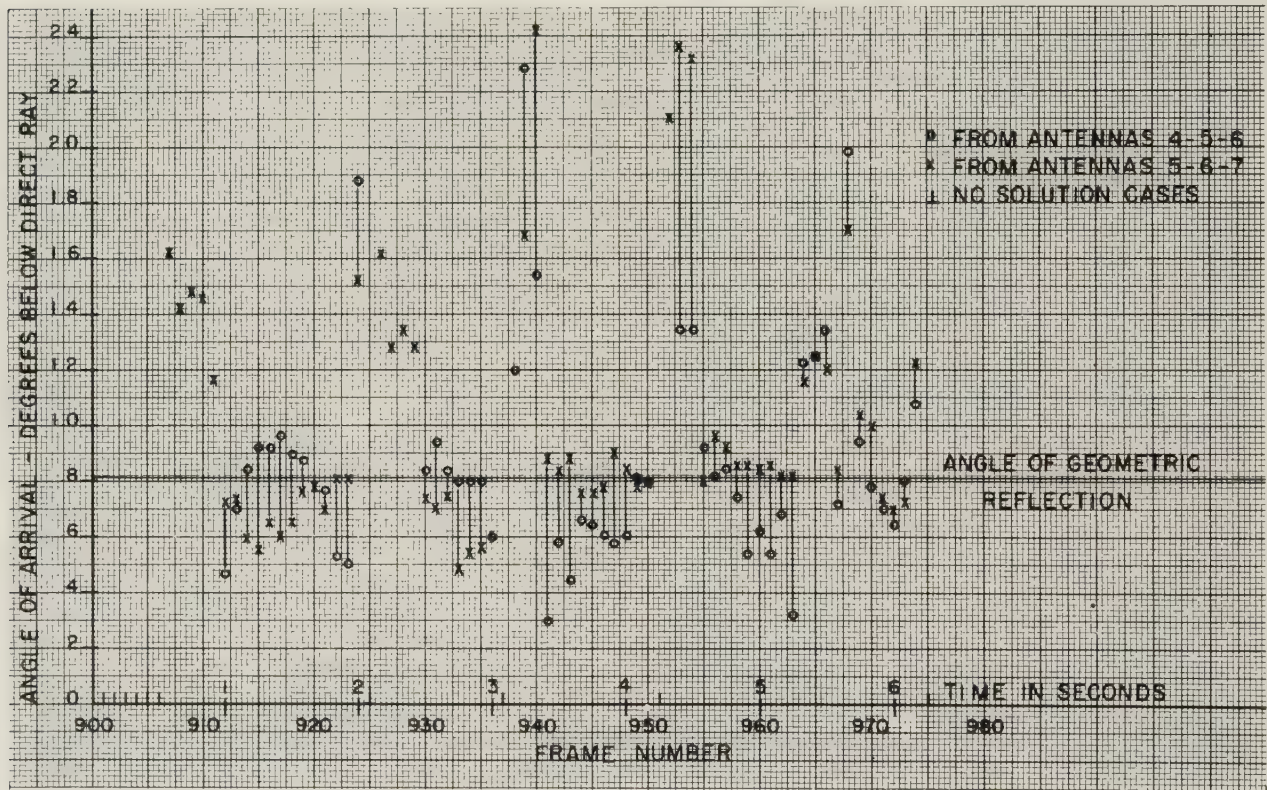
Fig. 11—Angle of arrival of variable component as a function of time.

Fig. 12 shows the distribution of the components of I, in-phase and 90 degrees out-of-phase with the fixed reflected component, plotted as a function of time and Fig. 13 shows the statistical distribution of these points plotted on Gaussian probability paper. These statistical distributions show that the in-phase and out-of-phase components can be represented approximately by straight lines on the Gaussian paper. The rms of the fluctuations for the in-phase components is found to be 0.2 of the direct signal. The out-of-phase fluctuations have an rms of 0.1. It is noted that the median of the I_y component is essentially zero while the median of
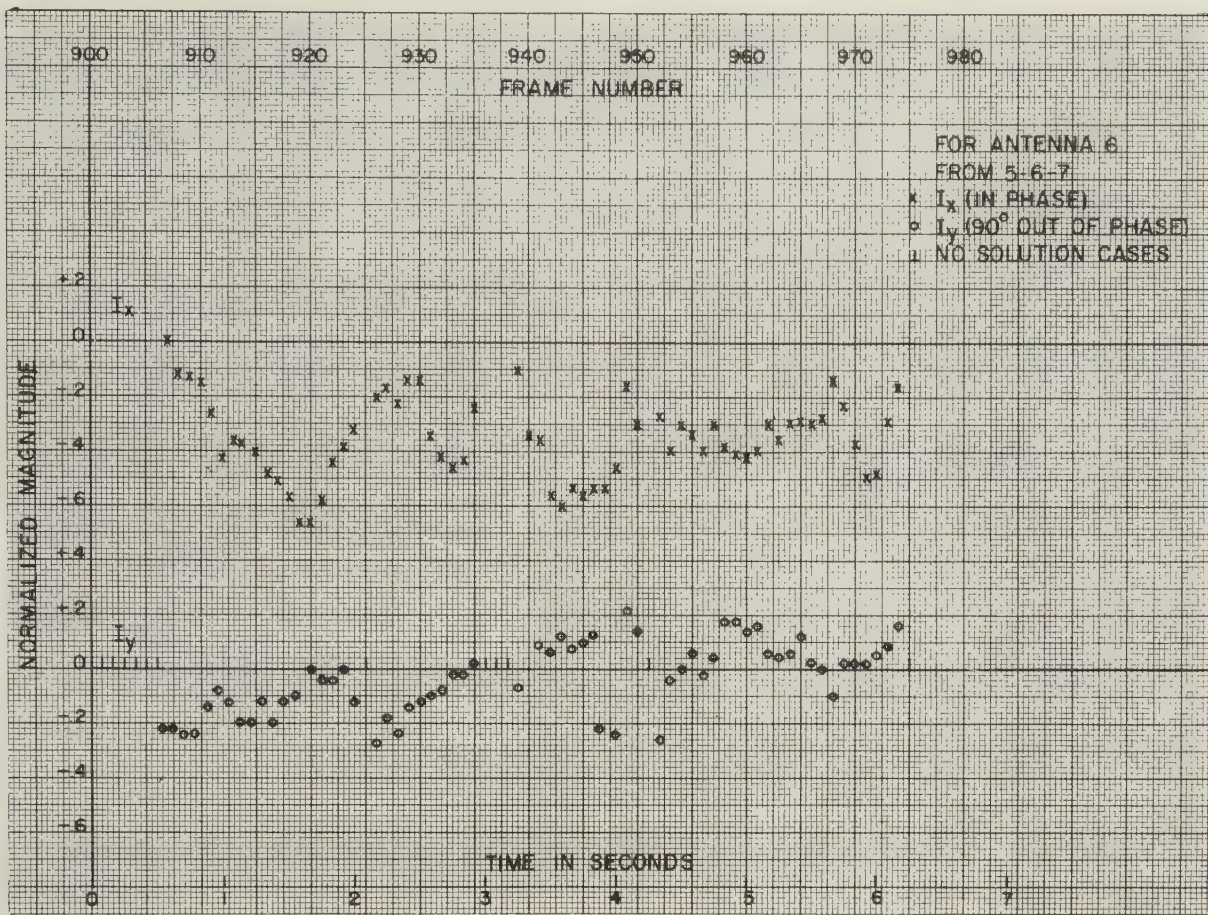
Fig. 12—Magnitude of I_x and I_y as a function of time.

the I_x component has a displacement from zero of -0.3 of the direct wave. The distribution of the total fluctuating term is essentially the same as the I_x component.

CONCLUSIONS

From the analysis described in this paper, it appears possible for this example to represent the fluctuation of the radio signal over a small height interval, as measured on a path over the rough surface of a small lake, as the vector sum of three plane waves. These waves are associated with the direct wave, a constant reflected wave and a varying reflected wave.

In the example studied, the component of the varying signal normal to the fixed reflected component was found to be Gaussian distributed about a median value of zero. The component of the varying signal in-phase with the fixed reflected component was found to be Gaussian distributed about the value of minus 0.3 of the direct wave.

Tests are under way for further measurement with the ten antenna system between drilling platforms in the Gulf of Mexico. Associated water wave measurements should make possible correlation studies of the nature of fluctuations and the characteristics of the water waves.
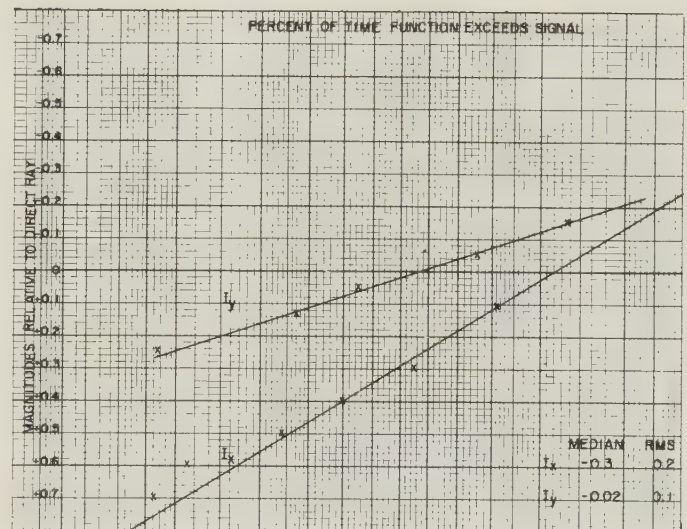
Fig. 13—Statistical distribution of I_x and I_y.

ACKNOWLEDGMENT

Measurements of the Phase of Signals Received over Transmission Paths with Electrical Lengths Varying as a Result of Atmospheric Turbulence

J. W. HERBSTREIT† AND M. C. THOMPSON†

Summary—A system for the measurement of the variations in effective lengths of radio propagation paths is described. The observed path-length instabilities are considered to be caused by the same atmospheric turbulence responsible for the existence of VHF and UHF signals far beyond the radio horizon. Preliminary results obtained on 172.8 mc and 1046 mc along a 3-1/2 mile path are reported. It is pointed out that measurements of this type should provide a powerful tool for the study of the size and intensity of the refractivity variations of the atmosphere giving rise to the observed phenomena.

Introduction

THE STABILITY of the phase of a signal propagated through the inhomogeneous atmosphere is important in considerations of the ultimate accuracy of radio direction finding systems and of the bandwidth capabilities of the medium. The spatial and temporal variations of the refractive index of the atmosphere[1-3] are now considered to explain the existence far beyond the radio horizon of useful vhf and uhf signal strengths much greater than are expected in the absence of such scattering. The magnitude and the spatial distribution of these inhomogeneities have been the subjects of both ground and airborne measurement by a number of investigators.[4, 5] Their results have been shown to provide a basis for the explanation, at least in a qualitative manner, of the transmission loss observations made over long propagation paths.[6, 7] The subject of this paper is the measurement of the effect of these inhomogeneities on the stability of the electrical path length (phase of arrival relative to the transmitted phase) of electromagnetic waves as they pass through the lower regions of the atmosphere. Theoretical aspects of the electrical path length stability have been treated in several papers.[8-10]

The National Bureau of Standards Cheyenne Mountain transmitting location 9,000 feet above sea level on the brink of a steep cliff overlooking the city of Colorado Springs (altitude 6,000 feet) was chosen for an initial study of the effects of the turbulent atmosphere on the phase of radio waves passing through the lower regions of the troposphere. Instrumentation has been developed to measure (1) the variations in the phase difference of the signals at the two ends of a single path and (2) simultaneously the variations in the phase difference between the signals arriving over the first path and those arriving over an adjacent path. In addition, instrumentation reported previously[1] has been developed to measure the very small changes in the amplitude of the received signal which occur within the radio horizon due either to scattering or to these phase variations. The use of these two types of instrumentation is expected to make possible the evaluation of the parameters important in the interpretation of scatter-type propagation mechanisms particularly when used in conjunction with microwave refractometers and other meteorological equipment under a variety of atmospheric turbulence conditions.
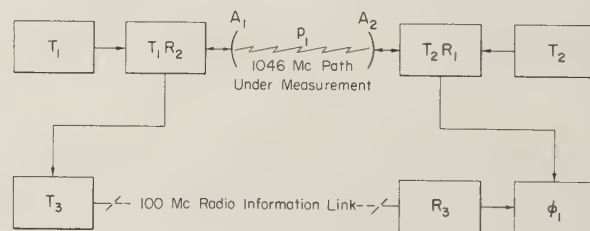
Fig. 1—Single-path measurement system.

Description of Phase Measurement System

The system for measurement of direct phase variations along a single path is shown in block diagram in Fig. 1. One transmitter on a nominal frequency of 1,046 mc is located at each terminal of the path and the frequencies are held to approximately two parts in 10^9

---

† National Bureau of Standards, Boulder, Colorado.

[1] A. P. Barsis, J. W. Herbstreit, and K. O. Hornberg, "Cheyenne Mountain Tropospheric Propagation Experiments," National Bureau of Standards Circular 554; January 3, 1955.

[2] H. G. Booker and W. E. Gordon, "A theory of radio scattering in the troposphere," Proc. IRE, vol. 38, pp. 401–412; April, 1950.

[3] H. Staras, "Scattering of electromagnetic energy in a randomly inhomogeneous atmosphere," *Jour. Appl. Phys.*, vol. 23, pp. 1152–1156; October, 1952.

[4] C. E. von Rosenberg, C. M. Crain, and A. W. Straiton, "Atmospheric Refractive-index Fluctuations as Recorded by an Airborne Microwave Refractometer," presented at URSI Spring Meeting, Washington, D. C.; also Elect. Engrg. Res. Lab., University of Texas, Report 6-01, 1953.

[5] H. E. Bussey and G. Birnbaum, "Measurement of variations in atmospheric refractive index with an airborne microwave refractometer," NBS *Jour. Res.*, vol. 51, pp. 171–178; October, 1953.

[6] W. E. Gordon, "Radio scattering in the troposphere," Proc. IRE, vol. 43, pp. 23–28; January, 1955.

[7] J. W. Herbstreit, K. A. Norton, P. L. Rice, and G. E. Schafer, "Radio wave scattering in tropospheric propagation," 1953 *Convention Record of the IRE*, Part 2, pp. 85–93.

[8] P. G. Bergman, "Propagation of radiation in a medium with random inhomogeneities," *Phys. Rev.*, vol. 70, pp. 486–492; October, 1946.

[9] V. A. Krasilnikov, "The effect of the coefficient of refraction in the atmosphere upon the propagation of ultrashort radio waves," *Izvestiya Akademic Nauk*, USSR, Seriya Geograficheskaya i Geofizich, vol. 13, pp. 33–57; 1949.

[10] A. D. Wheelon and R. B. Muchmore, "Line-of-Sight Propagation Phenomena," Ramo-Wooldridge Corp. Res. Lab. Rep., 1954; also paper presented at URSI Spring Meeting, Washington, D. C.; May, 1955.
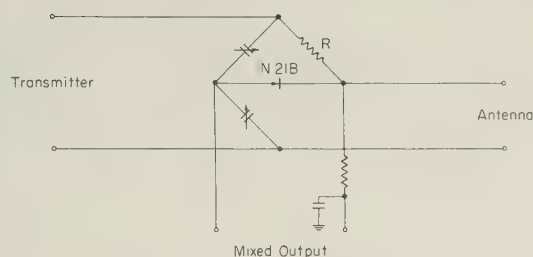
Fig. 2—Duplexer-mixer schematic.
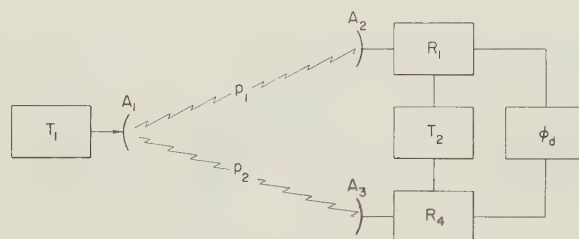
Fig. 3—Path-difference measurement system.
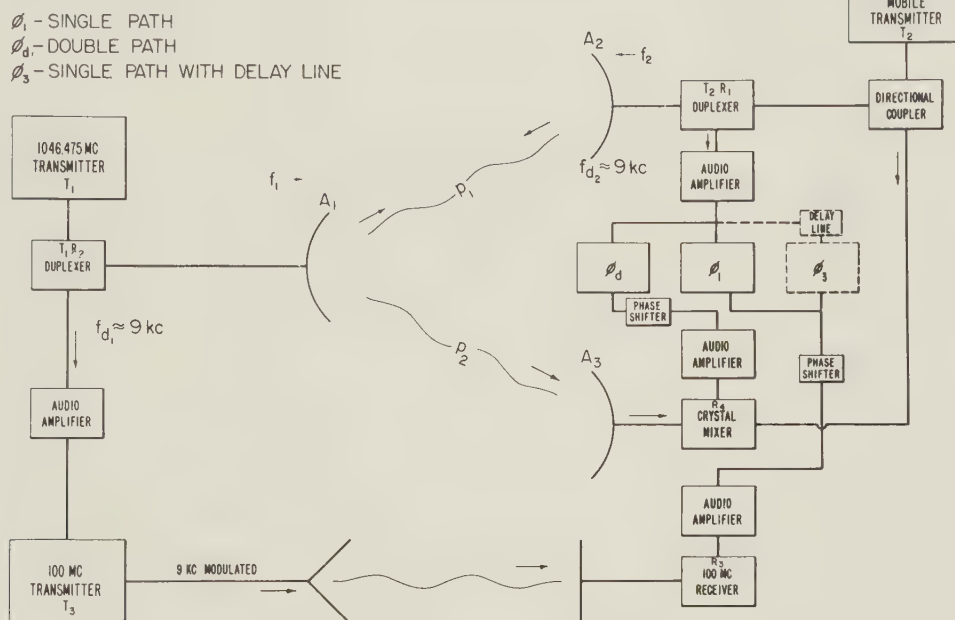
Fig. 4—Cheyenne Mountain 1,046-mc phase measurement system.

over a period of several hours. The frequencies of these transmitters, T_1 and T_2, are so chosen that their difference is approximately 9 kc. At each terminal a directional antenna is used together with a duplexing system such that each antenna simultaneously transmits one frequency and receives the other; the duplexer-crystal-mixer combinations are designated T_1R_2 and T_2R_1. These mixers are adaptations of coaxial bridge type directional couplers arranged with the antenna as one leg of an almost-balanced bridge, the crystal mixer being in the null portion of the circuit when observed from the transmitter, and in the high sensitivity portion when observed from the antenna. Thus only a very small portion of the high-level local transmitter energy appears across the mixer, a much larger portion being transmitted by the antenna; simultaneously a large portion of the energy received by the same antenna from the distant transmitter appears across the mixer. The circuit is illustrated in Fig. 2. A coaxial tuning stub is used as one capacitive arm of the bridge and also provides a dc return for the mixer.

The variations in the relative phase of the two 9 kc signals appearing at the output of the two mixers, one at each end of the path, are equal to twice the variations

in the electrical length of the path. In order to make the phase comparison, the 9 kc signal appearing at T_1R_2 is transmitted as a modulating signal on a 100-mc FM auxiliary transmitter to the other end of the path where it is received by a 100-mc auxiliary receiver and compared in phase by means of a recording electronic phase meter with the 9 kc signal detected in duplexer mixer T_2R_1.

In the two-path measurements illustrated in Fig. 3, the signal from the single transmitting system, T_1, is received on each of two receiving systems R_1 and R_4, some three miles distant from T_1. The two heterodyne receivers consisting of crystal mixers and audio amplifiers employ a common local oscillator which is actually the same transmitter, T_2, used in the single path measurements. The spacing of antennas A_2 and A_3 has been varied from 20 to 500 feet. The phases of the 9 kc signals from receivers R_1 and R_4 are compared in a separate electronic recording phase meter to record the variations in the difference in effective electrical path length of the two paths p_1 and p_2.

A block diagram of the complete system which integrates the two types of measurement (single-path and
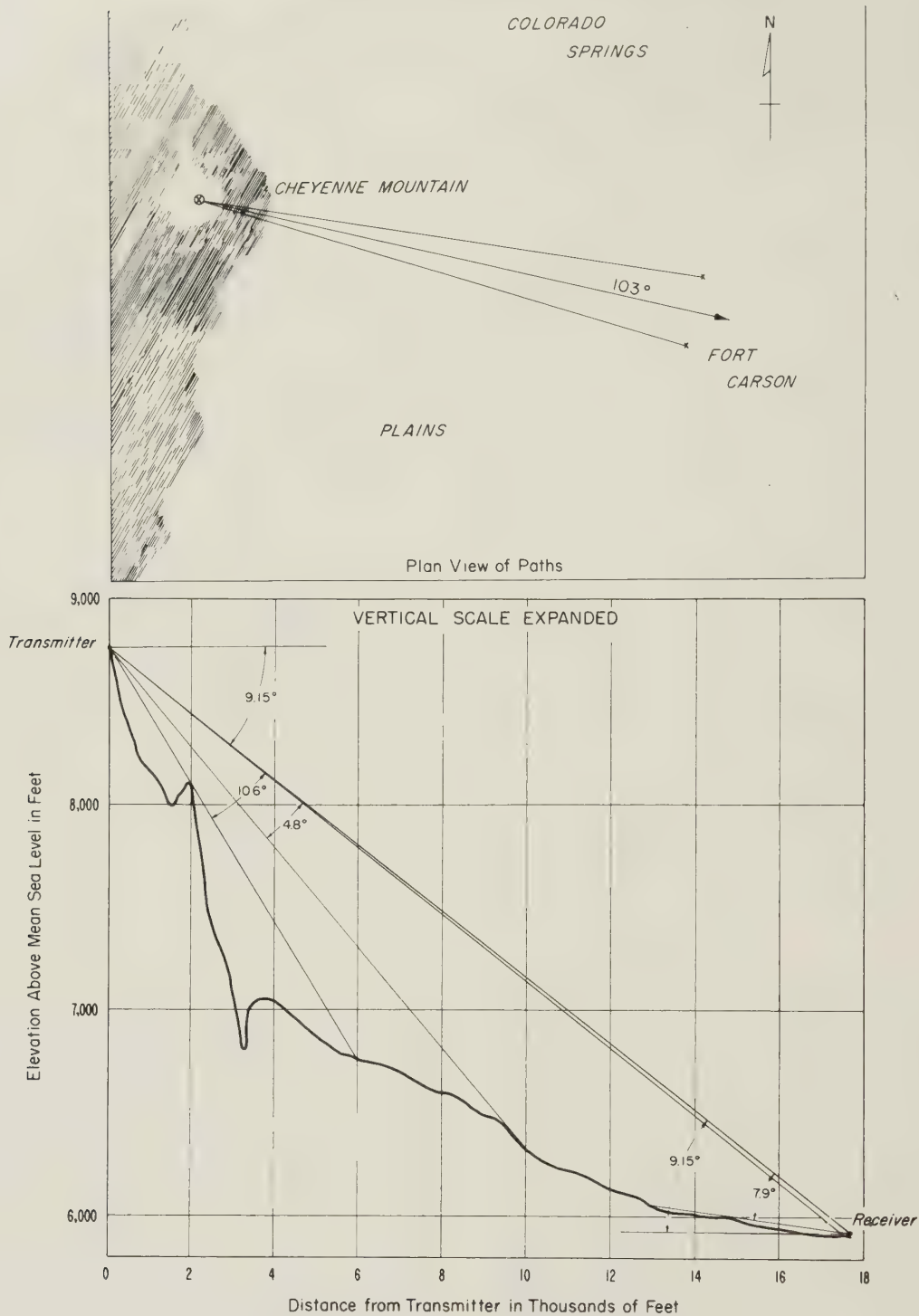
Fig. 5—Terrain profile Cheyenne Mountain summit to Fort Carson.

two-path) is shown in Fig. 4. An analysis of the system performance is given in the Appendix.

In order to study the effect of frequency on the effective electrical path-length difference an almost identical two-path system is in operation at 172.8 mc using equipment previously developed by E. Florman to determine the velocity of radio waves.[11] In addition, the University of Texas has operated a path-difference system at 9,375 mc for short periods on closely adjacent paths in conjunction with the NBS 172.8 and 1,046 mc measurements.

[11] E. Florman, "A measurement of the velocity of propagation of VHF radio waves at the surface of the earth," Journal of Research of Nat. Bureau of Standards Report vol. 54, No. 6, pp. 335-345; June, 1955.

PRELIMINARY RESULTS OF MEASUREMENTS

The National Bureau of Standards phase measurement system has been operated over a path from Cheyenne Mountain, Colorado, to Fort Carson for several periods of a few days each during the period from December, 1954, through April, 1955. The instrumentation has been undergoing continuous development to eliminate various sources of phase instability, the most serious of which has been the temperature instability of the transmission lines originally employed between the two spaced receiving antennas at the Fort Carson receiving site. Much of this phase instability has been eliminated by burying the transmission lines under ap-
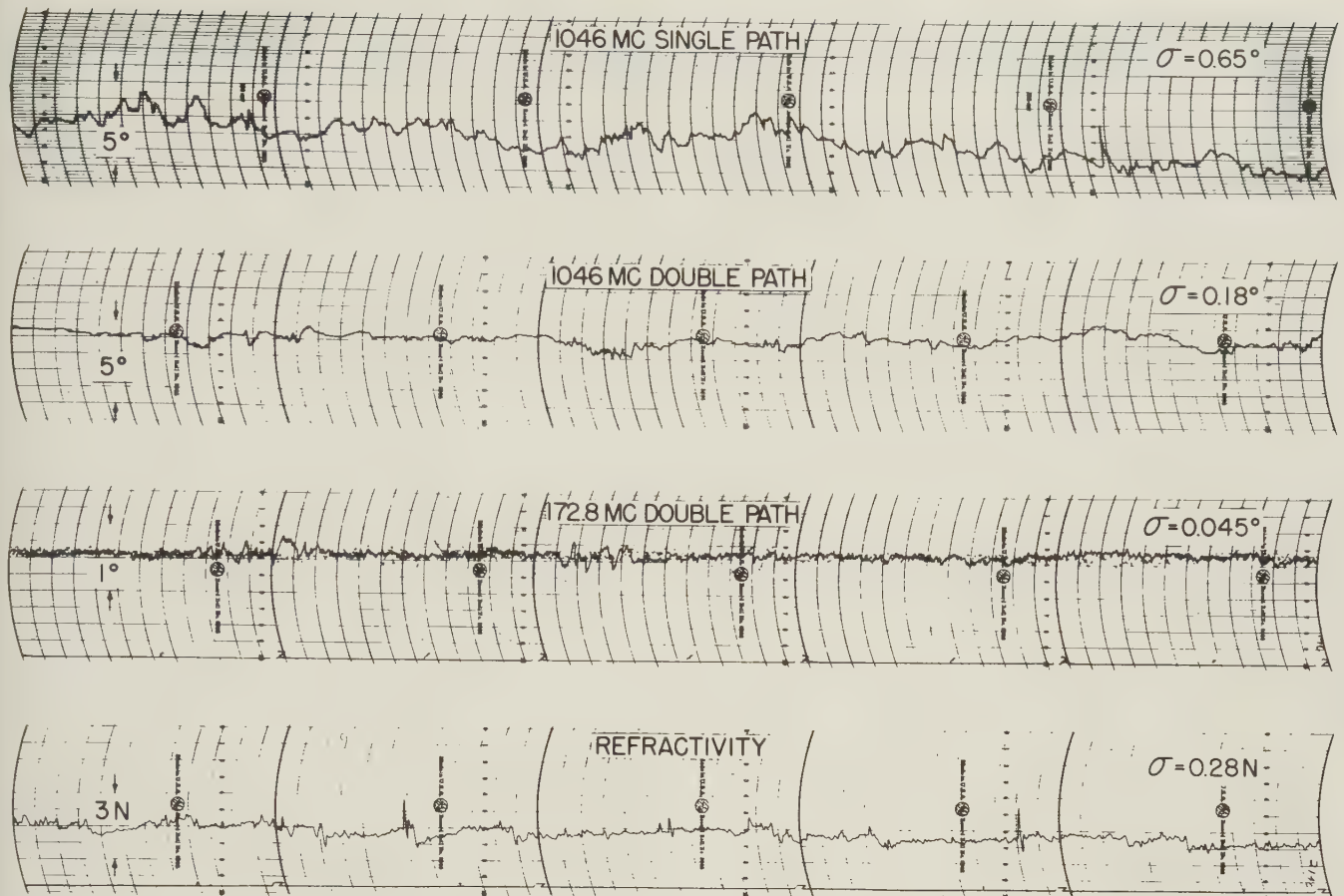
Fig. 6—Sample phase and refractometer recordings taken on Cheyenne Mountain to Fort Carson, Colorado path, March 15, 1955. (Antenna spacing of 500 feet for double-path measurements.)

proximately one foot of earth; however, steps have been taken to eliminate essentially all transmission lines from the 1,046 mc phase difference measurements by the installation of two separate single-path systems from which the path difference may be directly determined.

The geographical and terrain configuration of the path under study is shown in Fig. 5. The length of the direct path between the two sites is about 18,000 feet and the difference in elevation is approximately 2,900 feet. Simultaneous phase measurements have been made (1) over a single path at 1,046 mc, (2) at 1,046 mc on two paths with receiving antennas 500 feet apart, (3) at 172.8 mc on two paths with receiving antennas 500 feet apart, (4) at 9,375 mc on two paths with receiving antennas 500 feet apart. Simultaneously refractive index variations 25 feet above ground were measured at the Fort Carson site, and wind speed, wind direction, relative humidity, temperature and barometric pressure at both the upper and lower sites. Color photographs have been taken looking along the path at intervals of five seconds to record the existence of visible atmospheric phenomena such as clouds, dust, and precipitation.

In order to illustrate the type of data obtained, short samples of four kinds of records taken simultaneously with the NBS equipment are shown in Fig. 6; these records are believed to be representative of the results to be expected with equipment difficulties minimized.

Fig. 7 gives results of phase difference measurements
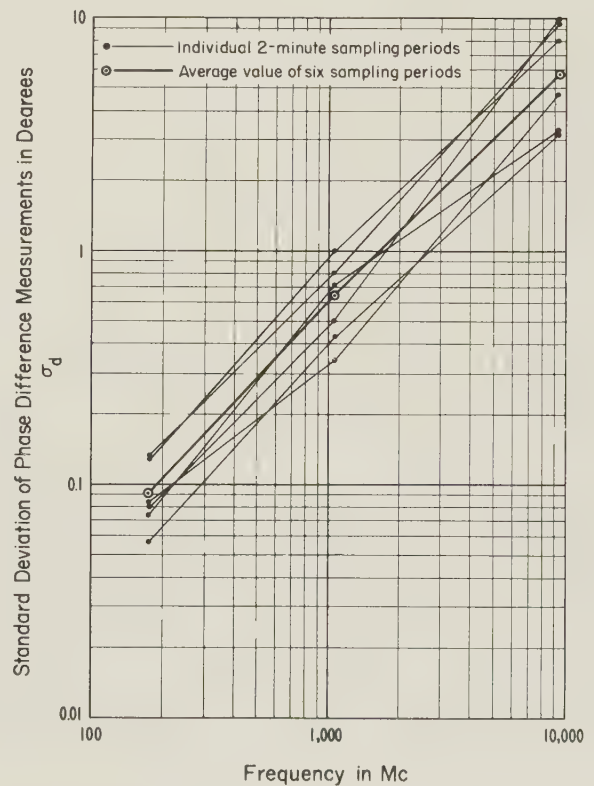
Fig. 7—Standard deviation of phase difference measurements vs radio frequency. Cheyenne Mountain to Fort Carson, Colorado path, 3:04–3:21 PM, March 18, 1955; antenna spacing 500 feet.

as a function of frequency for a short sample of data taken from a period when 9,375-mc data were being recorded simultaneously with those on frequencies of 172.8 and 1,046 mc; the 9,375-mc data were taken by the University of Texas. The standard deviation of phase differences, σ_d, was obtained for each of six 2-minute sampling periods mentioned above. At the time these data were taken, the antenna spacing on all three frequencies was nominally 500 feet. It is evident that σ_d was observed on this occasion to be very nearly directly proportional to the radio frequency in accordance with current theoretical predictions.[8-10] On other occasions this linear frequency dependence of σ_d was not found, but it is possible that equipment difficulties were responsible for these deviations from prediction. Further measurements with improved systems are planned for this summer with transmitters on Pikes Peak, and an opportunity will then be afforded for a further check of this frequency law and of other theoretical predictions.
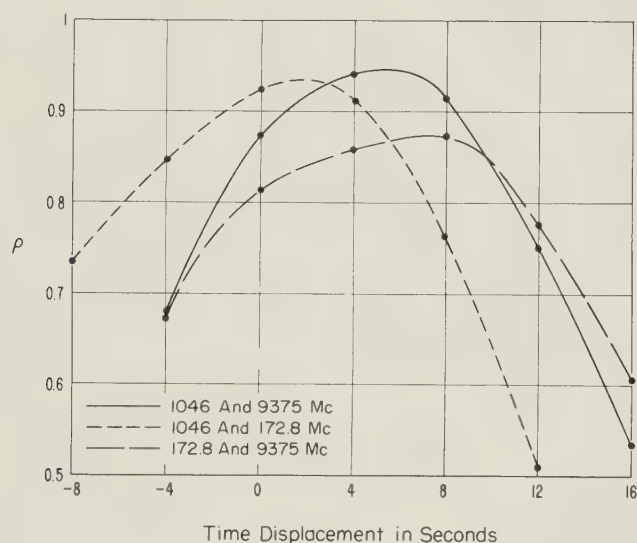
Fig. 8—Correlation coefficient of instantaneous phase difference vs time displacement. Cheyenne Mountain to Fort Carson, Colorado path, 3:16–3:18 PM, March 18, 1955.

In Fig. 8 the correlation coefficient is shown for one of the 2-minute sampling periods as a function of relative time displacement of the two records being compared. Going from south to north, the 1,046 mc, 172.8 mc, and 9,375 mc antenna systems were located on the baseline in that order, with approximately 50 feet separation between centers. The average wind speed was approximately 7 feet per second and essentially normal to the path. For the 50-foot system spacings, time delays of about 7 seconds in the correlation coefficients would be expected provided the atmospheric inhomogeneities which caused the changes in phase difference were moving with the average velocity of the wind. It is seen in Fig. 8 that the time delays corresponding to maximum correlation are of the correct sign, but their magnitudes are not in agreement with the above hypothesis.

A further variable which may be important in certain applications of systems depending upon the phase stability of propagation paths is the rate of change of electrical path length. Results of this type are also obtainable from these data.

Although the measurement of the meteorological factors considered important in producing variations in the effective electrical path length have been made at both ends of the propagation path under study, little or no correlation has been found thus far between the meteorological data and the radio data. Of course, it should be realized that a complete knowledge of the meteorological conditions in the volume of space along the path would be necessary to define adequately the propagation conditions. It is expected that refractometer measurements taken in an aircraft flying along the path will be more representative in this regard.

Conclusions

A system of measuring the variability of the effective length of the electrical path has been developed by the National Bureau of Standards. This system measures simultaneously (1) the variations in the effective electrical length of a single propagation path and (2) the difference in effective electrical lengths of propagation paths from a common transmitter to each of two spaced receiving antennas. These two sets of simultaneous phase measurements are expected to provide valuable data for the study of the nature of the refractivity variations of the atmosphere.

Wheelon and Muchmore[10] have derived theoretical expressions for the mean square phase variations of the direct wave which contain as parameters the frequency, the mean path length, a mean scale of turbulence and the mean square variation of the refractivity along the path. Wheelon, in a paper presented at this symposium[12] has derived expressions, depending principally upon the scale of turbulence, for the correlation of the effective electrical path lengths expected on two neighboring propagation paths. These theoretical derivations are based on various models of the turbulent atmosphere designed to explain scatter-type tropospheric propagation. Thus, our measurements of the variability of electrical path lengths provide a means for studying the applicability of these assumed models of the atmosphere and provide a radio measure of their parameters.

Appendix

The system for the measurement of the variations in the electrical path length essentially measures the variations in the time of propagation over the path. Variations in the time of propagation may be considered

[12] A. D. Wheelon, "Near-Field Corrections to Line-of-Sight Propagation," presented at the URSI Symposium on Electromagnetic Wave Theory, University of Michigan; June, 1955.

in terms of variations in the phase of arrival of the signal propagated with respect to a reference signal which has adequately stable phase characteristics. In the single-path system, signals are transmitted from both ends of the path. These same signals are also used as the phase reference signals after the time taken to traverse the path. This requires, in order to be able to measure the variations in transit time over the path, that the signal sources must have, over the transit time period, phase fluctuations which are small compared to the variations in phase of arrival introduced by the inhomogeneities of the propagation medium along the path. Since the signals are propagated in opposite directions, the transit times in two directions, even for the same time of transit, could conceivably be different particularly if changes in the medium occur nonuniformly along the path during the period of transit. There is a further complication introduced by the fact that the phase information for propagation in one of the directions must be transmitted to the point of phase comparison which requires an additional period of time.

The operation of the phase measurement system may be explained in the following manner: A signal of angular frequency ω_2 leaves transmitter T_2 at one end of the path and is propagated along this path toward the duplexer-mixer T_1R_2. After being in transit for the time interval Δt_1 this signal arrives at the opposite end of the path where it is received with the phase $\omega_2(1 + 3)\Delta t_1$ and is heterodyned with the signal of frequency $\omega_1(3)$ being transmitted from that end. The notation $\omega_2(1 + 3)$ indicates the frequency of transmitter T_2 at time $(t - \Delta t_1 - \Delta t_3)$ while $\omega_1(3)$ denotes the frequency of T_1 at time $(t - \Delta t_3)$; t is the time of phase measurement, and Δt_3 is the time required for the mixed signal to be transmitted over the 100-mc link. The value of the difference frequency at T_1R_2 is $[\omega_2(1 + 3) - \omega_1(3)]$. Similarly, a signal of angular frequency ω_1 leaves transmitter T_1 and traverses a path towards duplexer-mixer T_2R_1 where it arrives at some time Δt_2 later with the phase $\omega_1(2)\Delta t_2$ and is heterodyned with a signal from the local transmitter T_2, having an angular frequency of $\omega_2(0)$ at this time. The value of the difference frequency at T_2R_1 is $[\omega_2(0) - \omega_1(2)]$.

In general, since the structure of the atmosphere between the two sites is undergoing continuous change, the transit times for the signals traveling in opposite directions will not be equal to each other even when the signals are transmitted simultaneously from the opposite ends of the path. This difference in the transit times has been indicated above by using Δt_1 and Δt_2.

The actual measurement consists of obtaining the difference in phase between the two difference frequency signals just described. Since the system requires that the difference frequency signal generated at T_1R_2 be re-

turned to the other end of the path for comparison, this signal experiences the additional delay Δt_3 which is the time required for its transmission over the 100-mc link. Thus the signals responsible for the final phase comparison are those which were transmitted earlier by the interval $\Delta t_1 + \Delta t_3$ for the signal from transmitter T_2 and by Δt_2 for the signal from the transmitter T_1.

The results of the above discussion may be summarized in the following equation:

$$\phi_1 = \omega_2(1 + 3)(t - \Delta t_1 - \Delta t_3) - \omega_1(3)(t - \Delta t_3)$$
$$- [\omega_2(0)t - \omega_1(2)(t - \Delta t_2)]$$
$$= [\omega_2(1 + 3) - \omega_2(0)]t - \omega_1(2)\Delta t_2 - \omega_2(1 + 3)\Delta t_1$$
$$+ [\omega_1(3) - \omega_2(1 + 3)]\Delta t_3 + [\omega_1(2) - \omega_1(3)]t, \quad (1)$$

in which ϕ_1 is the instantaneous phase difference between the two difference frequency signals measured at the location of transmitter T_2 at reference time t; $\omega_2(0)$ is the angular frequency of transmitter T_2 at time t; $\omega_2(1 + 3)$ is the corresponding value of ω_2 at time $t - \Delta t_1 - \Delta t_3$; $\omega_1(2)$ is the angular frequency of transmitter T_1 at time $t - \Delta t_2$; and Δt_2 is the transit time for a signal arriving at T_2R_1 (from T_1) at time t.

From (1) it may be seen that in the case where $[\omega_1(3) - \omega_2(1 + 3)] \ll \omega_2$, the effect on ϕ_1 of variations in Δt_3 may be neglected. This equation also illustrates the way in which the frequency stabilities of the transmitters T_1 and T_2 directly influence the relation between the measurements and the propagation path behavior. In the system used for these measurements, $\Delta t_1 \overset{\sim}{\sim} \Delta t_2 \overset{\sim}{\sim} \Delta t_3 \overset{\sim}{\sim} 20$ μs, and $\omega_2 \overset{\sim}{\sim} \omega_1 \overset{\sim}{\sim} 2\pi \times 10^9$ radians per second. Under these conditions, $[\omega_2(1 + 3) - \omega_2(0)]t$ will introduce an error of 0.14 degrees as a result of the interval $\Delta t_1 + \Delta t_3$ when the stability of the frequency ω_2 is of the order of one part per 100 million. Since the interval $\Delta t_2 - \Delta t_3$ may be made essentially zero, the term of $[\omega_1(2) - \omega_1(3)]t$ will introduce negligible error. Thus the utility of the system is limited fundamentally by the "noise spectrum" of the frequency of transmitter T_2.

If the frequency drifts $[\omega_2(1 + 3) - \omega_2(0)]t$ and $[\omega_1(2) - \omega_1(3)]t$ are made sufficiently small, (1) becomes:

$$\phi_1 = -\omega_1(2)\Delta t_2 - \omega_2(1 + 3)\Delta t_1, \quad (2)$$

and the measured ϕ_1 will be a function of the propagation path during the two intervals Δt_2 and Δt_1, which do not coincide in time. A first-order correction may be made by inserting an additional time delay in the signal generated in duplexer-mixer T_2R_1 equal to the delay involved in the transmission over the 100 mc link (Δt_3). By introducing such a delay, an expression for ϕ_1 is obtained similar to (2) but in which the two transit times are essentially concurrent. This delay also has

the effect of decreasing the importance of the frequency stability of transmitter T_2, but increases proportionately the influence of the frequency stability of transmitter T_1. Preliminary experiments using such a delay line failed to show any significant differences in the measurements. However, due to equipment limitations simultaneous measurements of both delayed and undelayed signals were not made and it is felt that the preliminary tests are, therefore, inconclusive.

The development of an expression for the phase difference ϕ_d, over two paths of the same physical length is much simpler since the signals travel along both paths simultaneously. Thus

$$\phi_d = \omega_1(2)(t - \Delta t_2) - \omega_1(4)(t - \Delta t_4)$$
$$= [\omega_1(2) - \omega_1(4)]t + \omega_1(4)\Delta t_4 - \omega_1(2)\Delta t_2, \quad (3)$$

where $\omega_1(2)$ is the transmitted frequency at time $t - \Delta t_2$, $\omega_1(4)$ is the transmitted frequency at time $t - \Delta t_4$ and Δt_2 and Δt_4 are the transit times over paths p_1 and p_2 respectively. In this case it should be noted that there is only a very small second-order dependence on transmitter frequency stability to the extent that Δt_2 differs from Δt_4.

Conditions of Analogy Between the Propagation of Electromagnetic Waves and the Trajectories of Particles of Same Spin with Application to Rectifying Magnetrons

J. ORTUSI†

Summary—The object of this article is the study of the bi-univocal correspondence established by the Pauli principle between the internal energy of a particle and the frequency of the associated wave in media which are the seat of strong coupling between the particles and when their spins are in a favored direction.

In Section I, the determinantal forms of antisymmetrical wave functions are investigated, these being valid both for crystals and for electronic plasmas. It is shown that, starting from this determinant, two complementary series of wave functions can be constructed. Depending on the internal energy, two types of complementary particles are thus obtained: (1) free electrons associated with real waves, and (2) holes associated with evanescent waves.

In Section II, a study is made of the mathematical analogy between the Schrödinger equation and the tropospheric propagation equation. It is shown that the potential energy can be assimilated to the refraction modulus and that the group velocity of the propagation around the earth can be assimilated to the group velocity of the complementary particle.

By a very simple correspondence, the real modes of propagation predict the formation of holes while the imaginary modes of propagation predict the formation of free electrons. A special study is made of the analogy between the index barriers of the inversion layers and the potential barriers of the barrier layers. This analogy enables the existence to be predicted of purely electronic barrier layers without the need for any material support.

In Section III, the rectification and photoconduction properties of these electronic barrier layers in magnetrons and in traveling wave magnetron detectors are considered. Their analogies with and differences from the barrier layers of p-n junctions are examined. Finally, in the conclusion, the advantages and description of radar detection arrangements devised, on these principles, by the Compagnie Générale de T.S.F. in Paris, are set out.

INTRODUCTION

As is well known, particles are divided into two main classes:

Particles not coupled together, known as bosons, calculated by the Bose-Einstein statistics.

Particles coupled together, known as fermions, calculated by the Fermi statistics.

In the latter case, the particles obey the exclusion principle. Only two particles, of different spins, may occupy the same energy level. However, in numerous examples in physics, crystalline or noncrystalline media are considered, which contain electrons whose proper magnetic moment has a favored direction; e.g., in the presence of external magnetic fields. The study of such media is then conditioned by the study of the electronic media in which spins are identical.

In such media, the exclusion principle establishes a *bi-univocal correspondence* between the particle energy ϵ and the frequency of the wave ν associated with this particle. This bi-univocal correspondence allows the decomposition, by the phase integral method, of the

wave function of N particles into its elementary components each defined by the frequency of the elementary wave. This frequency will represent, by the Einstein relation $\epsilon = h\nu$, the energy of the particle corresponding to this elementary wave and the amplitude of the elementary wave will represent the density of particles having this energy.

In what follows, we shall use this bi-univocal correspondence in order to find analogies with the studies of tropospheric propagation.

Both the crystalline media, in which the wave functions are periodic and are represented by space harmonics analogous to Fourier series, and electronic media, with high space charge causing particles to be coupled together, will be considered; the wave functions in the latter media are then nonperiodic and are represented by space harmonics analogous to Fourier integrals.

In Section I, we shall examine the general form of the wave function in such media and establish that they contain *particles complementary* to the free particles, associated with evanescent waves. In section II, we shall examine the mathematical conditions enabling analogies between the wave functions in these media and tropospheric propagation to be predicted. In Section III, these conditions will be utilized in order to predict the existence of *electronic barrier layers*, even in the vacuum, without any material support, and to study their properties in detecting magnetrons.

I. GENERAL FORM OF THE WAVE FUNCTION IN MEDIA WITH COUPLED PARTICLES

Anti-Symmetric Wave Functions

Let us consider N identical particles; *e.g.*, N electrons in the same sublayer, in a system of N atoms first considered as being infinitely far apart. Plotting as ordinates the energy of each electron (energy level diagram) and as abscissae the inter atomic distance, it is of course found that the N levels then merge in a single horizontal line.

But, when the inter atomic distance is reduced, the energy levels are distributed over N values which are further apart as the distance d diminishes. These N values constitute a band of energy levels.

This theorem, which constitutes Pauli's exclusion principle, can very simply be interpreted physically. Each electron can be assimilated to an elementary resonator with a resonance frequency ν and a stored energy ϵ connected with ν by the Einstein relation $\epsilon = h\nu$.

When the interatomic distance is very large, there is no coupling between the resonators which then all have

---

† Compagnie Générale de T.S.F., Paris, France.

the same resonance frequency, and hence the same energy.

When the interatomic distance diminishes, a certain degree of coupling arises between all resonators, because of the interpenetration of the associated waves.

Because of this coupling, each resonator has its resonance frequency modified. By analogy with the theory of coupled circuits, it is known that there are then N different resonance frequencies, and, hence, N different values of the energy—with greater spacing as the coupling increases.

This theorem is expressed by quantum mechanics. The possibility of distinguishing between two particles, when the spectral distributions follow a Gaussian probability law, is then defined, on the basis of the principle of uncertainty, by the possibility of expressing the wave function in the form of a normal probability law.

$$\begin{cases} |u_n|^2 = \dfrac{1}{a\sqrt{\pi}}\, e^{-\frac{(x-\bar{x})^2}{a^2}} \\ \quad\text{or} \\ u_n = \dfrac{1}{\sqrt{a\sqrt{\pi}}} \cdot e^{-\frac{(x-\nu_0 t)^2}{2a^2}} \cdot e^{-2i\pi\nu\,(t-x/V)} \end{cases} \quad (1)$$

where a is the precision parameter of the normal law; $(a \equiv \Delta x)$, $\bar{x}$ is the mean value of the position x $(\bar{x} = \nu_0 t)$, V is phase velocity, and ν_0 is group velocity connected with V by the relation

$$mV\nu_0 = mc^2 = \epsilon = h\nu.$$

This possibility is excluded within an atom in which the distinction between the various electrons only expresses the existence of the energy levels of the various layers. It exists between the electrons of different atoms, well separated from each other. Two identical particles, i.e., particles capable of being interchanged without physical change in the phase configuration, can then be distinguished.

In consequence, let ψ_1, ψ_2, $\cdots$ ψ_j, $\cdots$ ψ_N be the respective wave functions of N identical particles, of same spin, in the corresponding space of each.

Let us designate by ϕ_N the normalized wave function corresponding to the simultaneous presence of the N particles. When there is no interaction between the particles, the probability of simultaneous presence in each space is the product of the probabilities of presence of each particles in the corresponding space:

$$\phi_N = K\psi_1\psi_2 \cdots \psi_j \cdots \psi_N, \quad (2)$$

K being a normalization coefficient.

Such a wave function is said to be *symmetrical* because the interchange of two particles leaves the function ϕ_N unchanged.

Conversely, when a certain degree of interaction exists between identical particles, it is necessary to define the wave functions associated with the presence of a particle in the various spaces attributed respectively to the other $N - 1$ particles.

These coupling functions are represented by the notation $\psi_j(k)$ with $j \neq k$.

The notation $\psi_j(j)$ represents the wave function corresponding to the proper space of the particle of order j.

The expression of the wave function ϕ_N, when the coupling between the particles is suitable, is given in the form of a *determinant*:

$$\phi_N = K' \cdot \begin{vmatrix} \psi_1(1) \cdots \psi_1(k) \cdots \psi_1(j) \cdots \psi_1(N) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \psi_k(1) \cdots \psi_k(k) \cdots \psi_k(j) \cdots \psi_k(N) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \psi_j(1) \cdots \psi_j(k) \cdots \psi_j(j) \cdots \psi_j(N) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \psi_N(1) \cdots \psi_N(k) \cdots \psi_N(j) \cdots \psi_N(N) \end{vmatrix}. \quad (3)$$

K' also designating a normalization coefficient

Such a wave function is said to be *anti-symmetrical* because the exchange of two particles modifies the sign of ϕ_N. It is clear that the exclusion principle is an immediate consequence of the determinantal form of the anti-symmetrical functions. If the particles of order j and k have identical wave functions, including the coupling wave functions, the determinant then has two identical rows and is therefore zero.

Complementary Particles

Let us also consider a number N of particles of same spin obeying the exclusion principle. We have just seen the determinantal form of the wave function associated with the simultaneous presence of N particles.

An important problem is that of establishing the new form of the wave function when one particle is dissociated from the other $N - 1$ particles.

The dissociation of this particle implies that the associated wave function φ is to be given the particular form u_n of (1), representing the wave packets.

Such a particle is said to be *free*. It is characterized by its velocity $\vec{v}$, its momentum $\vec{p}$ and its energy ϵ. These magnitudes are connected by the relation

$$\vec{v} = \operatorname{grad}_p \epsilon(\vec{p}), \quad (4)$$

in which the gradient operator is applied to the representative space of vector $\vec{p}$.

Fig. 1—Variation of energy and velocity with crystal momentum within a Brillouin zone.

In a crystal, $\vec{p}$ is called crystal momentum. The function $\epsilon(\vec{p})$ is represented in Fig. 1 within the Brillouin zone.

For the cross-hatched part of the curves, the following relations can be written:

$$\begin{cases} \vec{p} = \vec{p} = m\vec{v} \\ \epsilon = \epsilon_o + \dfrac{p^2}{2m} = \epsilon_o + \tfrac{1}{2}mv^2 = \epsilon_o + \epsilon_c, \end{cases} \quad (5)$$

where ϵ_o is a universal constant and ϵ_c is the positive kinetic energy of the free particle.

Also, the wave function has, for a proper solution, within the crystal, a Bloch function:

$$\vec{\psi}_j = v(\vec{r})e^{2i\pi(vt + i\vec{p}\cdot\vec{r}/h)} \quad (6)$$

$\vec{r}$ designating the position vector, and $v(r)$ a periodic function with respect to the crystal distance d.

The Bloch functions, being eigenfunctions, can, in accordance with the well-known properties of eigenfunctions, be employed to represent the function associated with the free particle [(1)] in the form of a series.

$$u_n(\vec{r}, t) = \sum_{j=1}^{j=N} a_j(t)\cdot\psi_j. \quad (7)$$

Similarly, these functions can determine, in accordance with the process examined in the preceding paragraph, the wave function $u_p(\vec{r}, t)$ associated with the simultaneous presence of the $N - 1$ remaining particles.

If we consider the determinant of (3), where the terms ψ_j are Bloch functions, [6], it is clear that each minor term M_j of this determinant represents an antisymmetrical eigenfunction which is a solution of the Schroedinger equation associated to the presence of $N - 1$ particles. It is then possible to establish the solution[1] from a series of these eigenfunctions:

$$u_p(\vec{r}, t) = \sum_{j=1}^{j=N} b_j(t)\cdot M_j. \quad (8)$$

Calculation further shows that b_j is no other than the complex conjugate quantity of a_j:

$$b_j = a_j^* \quad (9)$$

Also, because of the possibility of being able to add the eigenvalues corresponding to the eigenfunctions constituting the determinant, the eigenvalue of the energy of the eigenfunction M_j is equal to

$$\epsilon_{Mj} = \epsilon_1 - \epsilon_{\psi j}, \quad (10)$$

ϵ_1 designates the constant total energy, and $\epsilon_{\psi j}$ is the eigenvalue of the energy corresponding to the eigenfunction ψ_j.

By definition, we shall designate the eigenfunctions ψ_j and M_j by the term *complementary eigenfunctions*.

The particles whose wave functions are given by the series (7) and (8) will be designated as *complementary particles*.

We have already examined the case of the particles defined by series (7). These particles are free particles (electrons or ions). The particles defined by series (8) are called *holes* in the case when the N particles are

electrons, and *deficiency ions* in the case when the N particles are ions.

It is easy, using (9) and (10), to show that holes and deficiency ions have the following complementary characteristics respectively for free electrons and free ions. Their masses are of the same order of magnitude; their charges are of opposite sign. Eq. (4) is applicable to the complementary particles. But the (5) becomes:

$$\begin{cases} \vec{p} = im\vec{v} \\ \epsilon = \epsilon_o - \dfrac{p^2}{2m} = \epsilon_o - \tfrac{1}{2}mv^2 = \epsilon_o + \epsilon_c. \end{cases} \quad (5 \text{ bis})$$

Similarly, the form of Bloch function (6) becomes

$$\psi_p = u(\vec{r})e^{\frac{2i\pi}{h}\vec{p}\cdot\vec{r}}\cdot e^{2i\pi vt}. \quad (6 \text{ bis})$$

The value of $\vec{p}$ being purely imaginary, the wave function ψ_p represents a *pseudo-periodic evanescent wave*, and ϵ_c is < 0. Similarly, the representative form of the evanescent wave packet associated with a hole circulating at a velocity $\vec{v}$, when the spectral distribution also follows a Gaussian probability law, is given by equation

$$u_p = \frac{e^{\frac{-2\pi mvx}{h}}\cdot e^{-\left|\frac{x^2 - v^2 t^2}{2a^2}\right|}\cdot e^{-2i\pi vt}}{\sqrt{a\sqrt{\pi}}} \quad (1 \text{ bis})$$

Further, (10) shows that two particles of kinetic energy ϵ_c and ϵ_{c1} are complementary if the relation $\epsilon_c + \epsilon_{c1} = 0$ is verified, and so velocities are the same.

II. Conditions of Analogy Between the Trajectories of Particles of Same Spin and Tropospheric Propagation

Decomposition into Real Modes and Evanescent Modes of the Tropospheric Field

It is well known that the field of a tropospheric propagation, according to the classical calculation method, is given by a series of real or of evanescent modes according to the shape of the index profile curve.
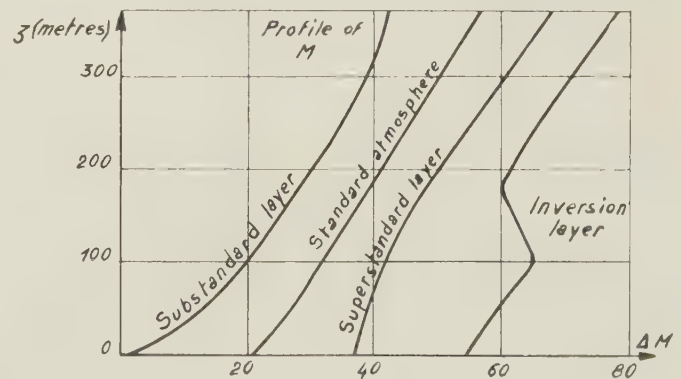
Fig. 2—Variation of modified index $N = n + z/R$ or of $M = (N - 1)\cdot10^6$.

Fig. 2 shows curves of index profile giving, as a function of the height z, the modified index $N = n(z) + z/R$ or the modulus of refraction $M = (N - 1)10^6$

The solution giving the field at a point P at a distance d from the transmitter and at height z, is given by the

---

[1] Shockley, "Electrons and Holes in Semi-Conductors," Bell Labs. Ser., p. 442;

well-known Sommerfeld integral. The solution of this integral in the form of a series of modes constitutes the Fundamental Theorem. According to this theorem, the mode of order n of the vertical component of the hertzian vector (and, by differentiation, the electromagnetic field) is given, at great distance, by the formula

$$U_n = \frac{e^{-i\mu_n d}}{\sqrt{d}} \cdot v(z, \mu_n).$$

The function v represents the gain in height and its proper values determine the real or evanescent nature of the mode. It is given, with appropriate boundary conditions, by the solution of the differential equation

$$v'' + (k_o{}^2 N_{(z)}{}^2 - \mu_n{}^2)v = 0 \quad \text{with} \quad k_o = \frac{2\pi}{\lambda_o}, \tag{11}$$

where λ_o is the wavelength. When μ_n is real, the wave is real. The mode is called a *Gamow mode*. When μ_n is imaginary, the wave is evanescent. The mode is called an *Eckersley mode*.

In the case of the standard atmosphere, the index N varies linearly as a function of z and it can be shown that, at a sufficient distance, only the Eckersley modes are possible. This happens whatever the dielectric constant and conductivity of the ground.

Conversely, in the case of an inversion layer (the right curve of fig. 2), it can be shown that there exist Gamow modes created inside the inversion layer which acts like a waveguide. These modes are fairly independent of the dielectric constant and conductivity of the ground.

Mathematical Conditions of Analogies

In this Section, we shall show the profound analogy which exists between tropospheric propagation and the propagation of the wave associated with a particle when the potential energy depends only on the co-ordinate z, $[\text{II} \equiv \text{II} (z)]$

An elementary particle is defined by:
  the position O,
  the velocity $\vec{v}$.
  the rest mass m_o of electrons and holes,
  the internal energy $\epsilon_n = \epsilon_o + \epsilon_c$ in which
    ϵ_o is a universal constant equal to $5.10^5 \, eV$,
    ϵ_c is the kinetic energy, positive for an electron, negative for a hole,

ϵ_c is connected with v by the relation common to both elementary particles $|\epsilon_c| = \frac{1}{2}m_o v^2$.

This relation, in the case of solids, remains valid for the centre of the Brillouin zone, corresponding to energy levels close to the Fermi level.

Schroedinger Equation. The trajectory of the particle assumed to be on an axis O_z, the associated wave, in cylindrical coordinates z and r, will be represented by the expression

$$\varphi_1(r, z)e^{i\omega_n t} \equiv \varphi(r, z)e^{i(\omega_n t - k_s z)} ;$$

in which

$$w_n = k_n c = \frac{2\pi c}{\lambda_n}$$

is the angular frequency,

$$k_z = \frac{2\pi}{\lambda_z}$$

is the propagation constant.

When k_z is very small compared to k_n, the function φ_1 must satisfy the Schroedinger equation

$$\Delta\varphi_1 + \frac{8\pi^2 m_n}{h^2} (\epsilon_n - \text{II}_{(z)})\varphi_1 = 0 \tag{12}$$

According to the bi-univocal correspondence of energy and frequency in particles of same spin, to any mode of the function $\varphi_{1\,r}$, there should correspond a particle, with the two following reservations:

1. The eigenvalue ϵ_n must be a *positive* real number equal to the internal energy of the particle.

2. To correspond to the solution of the stated problem, the associated wave must satisfy the uncertainty relations relevant to that problem.

In that case, the wave associated to the particle is similar to a wave carrying a signal having a Gaussian spectral distribution with Δ_ν as parameter and propagated in a circular waveguide of O_z axis whose radius Δ_r is of the order of the wavelength at rest.

The energy is exponentially attentuated outside the guide. Inside it, there is a constant energy E_o. This is the case of the particle at rest, when the wave length λ_z is infinite. When the energy is modified along a certain part of the guide having a Gaussian distribution with Δ_z as parameter, one can speak of the circulation in this part of a particle.

When the energy E_o is increased, it is an electron analogous of a wave upon the sea.

When the energy E_o is decreased, it is a hole analogous of a vortex under the level of the sea.

In cylindrical coordinates, the diameter $2\Delta_r$ of the guide is given by the uncertainty relation and its minimum is equal to the wavelength at rest.

We have therefore:

$$2\Delta r \text{ (min.)} = \frac{h}{m_o c} = \lambda_o = \frac{2\pi}{k_o} ; \tag{13}$$

λ_o wavelength associated with the particle at rest.

To each eigenvalue ϵ_n there will correspond, under condition (13), a particle of velocity v_n and mass m_n bound by the classical relations

$$\begin{cases} \epsilon_n = m_n c^2 \\ |\epsilon_c| = \frac{1}{2}m_o v_n{}^2 \\ \epsilon_o = m_o c^2 \end{cases} \begin{cases} \lambda_n = \dfrac{c}{v_n} = \dfrac{h}{m_n c} = \dfrac{2\pi}{k_n} \\[2mm] |\lambda_z| = \dfrac{|V_n|}{v_n} = \dfrac{h}{m_n v_n} = \dfrac{2\pi}{k_z} \\[2mm] \lambda_o = \dfrac{c}{v_o} = \dfrac{h}{m_o c} = \dfrac{2\pi}{k_o} \end{cases} \tag{14}$$

with the relativity condition

$$\epsilon_n = \epsilon_o + \epsilon_c$$

which gives

$$k_n{}^2 = k_o{}^2 + k_z{}^2.$$

When ϵ_n is greater than ϵ_o, $k_z{}^2$ is positive; the mode is real. The particle associated to the eigenvalue ϵ_n of (12) is a *free electron* ($\epsilon_c > 0$)

When ϵ_n is less than ϵ_o, $k_z{}^2$ is negative; the mode is evanescent. The particle associated to the eigenvalue ϵ_n of (12) is a *hole* ($\epsilon_c < 0$), complement of the free electron of same velocity v_n.

Conditions of Analogy Between the Equations. Let us take up (12), separating the variables by making

$$\varphi_1(r, z) = \zeta(r)\psi(z)e^{-ik_z z}.$$

The Schroedinger equation is reducible to the two differential equations:

$$\zeta'' + \frac{\zeta'}{r} + \sigma^2\zeta = 0, \qquad (15)$$

$$\psi'' + \frac{8\pi^2 m_n}{h^2}(\epsilon_n - \mathrm{II}_{(z)})\psi - (\sigma^2 + k_z{}^2)\psi = 0; \quad (16)$$

σ designating a constant.

The first equation is integrable and has the solution

$$\zeta = J_o(\sigma r].$$

In actual fact, when considering the wave associated with a particle in the presence of P particles, it can be shown that the Bessel function $J_o(\sigma r)$ must be replaced by the function

$$\zeta_1 = \frac{J_o[\sigma(r - r_p)]}{J_o[\sigma r_p]} \cdot e^{\frac{\sigma^2 r^2}{2}}$$

in which σr_p designates the Pth root of the function J_1.

In order that this function may represent a solution connected with determination of a trajectory, we must, in addition, have Δr (min) equal to the distance obtained by the first two zeros of the function ζ_1.

If P is fairly large, it is readily seen that

$$\sigma \Delta r \text{ (min)} = \pi.$$

Taking into account the value of Δr supplied by (13), we have

$$\sigma = k_o.$$

Eq. (16) becomes, for this value of σ,

$$\psi'' + \frac{2k_n k_o}{\epsilon_o}[\epsilon_n - \mathrm{II}(z)]\psi - k_n{}^2\psi = 0. \qquad (17)$$

By a limited development, assuming $k_z \ll k_o$, and $\mathrm{II}_{(z)} \ll \epsilon_o$, we have

$$\begin{cases} k_n = k_o + \dfrac{k_z{}^2}{2k_o} \\ \epsilon_n = \epsilon_o\left(1 + \dfrac{k_z{}^2}{2k_o{}^2}\right) \end{cases}$$

and

$$\psi'' + \left[k_o{}^2\left(1 - \frac{2\mathrm{II}_{(z)}}{\epsilon_o}\right) + k_z{}^2\right]\psi = 0 \qquad (18)$$

Comparing (18) with the equation for tropospheric propagation (11),

$$\psi'' + [k_o{}^2 N(z)^2 - \mu_n{}^2]\psi = 0, \qquad (11)$$

it is seen that (18) and (11) can be identical, when $N_{(z)}$ is close to unity, provided we set

$$N(z) - 1 = M(z)10^{-6} \equiv -\frac{\mathrm{II}(z)}{\epsilon_o}$$

$$\mu_n{}^2 \equiv -k_z{}^2.$$

As will be readily seen, these relations are remarkably simple. Thus, apart from boundary conditions, a propagation problem can be assimilated to that of the trajectory of a particle with a given potential energy $\mathrm{II}_{(z)}$, when the following conditions are fulfilled:

1. *The potential energy $\mathrm{II}(z)$ is assimilated to the modulus of refraction $M(z)$.*[2]

2. *The group velocity of the propagation along the earth is assimilated to the group velocity of the complementary particle along its trajectory.*

But the solutions can be identical only if the boundary conditions are the same or if they have no great influence on the form of the solutions.

In the case of free particles, the boundary conditions are always verified, and there are no eigenvalues for ϵ_c.

In the case of bound particles, the boundary conditions require that ϵ_c as in the case of μ_n shall take a group of eigenvalues ϵ_{cn}.

Application of the Analogies between the Two Studies

The application of this analogy consists of certain predictions relative to trajectories and, consequently, to transient phenomena and the mean free path of the particles bound to the function $\psi_{(z)}$. In general, and within boundary conditions,

The *Gamow modes*, corresponding to real modes, will be assimilated to waves associated to the relevant complementary particles, i.e. *to holes*;

The *Eckersley modes*, corresponding to evanescent modes, will be assimilated to waves associated to the relevant complementary particles; i.e., *to electrons*.

The importance of the value of the index gradient stresses the important role played by the potential energy gradient, proportional to the electric field

$$-\vec{E} = \frac{1}{e}\,\mathrm{grad}\,\mathrm{II}(z) = \mathrm{grad}\,V(z);$$

e designating the charge, negative for the electron and positive for the hole. We shall consider exclusively two particular cases of assimilation:

1. *Assimilation of the Standard Atmosphere*

This can be assimilated to a linearly variable potential, hence a relatively weak constant field. This is a very general case where the potential follows Ohm's law.

It has been stressed that the only possible modes of propagation are then the Eckersley modes, this being so for any boundary condition.

There results, by assimilation, the following principle:

When, in a solid or in a vacuum, the electric field remains constant, only the electron particles can circu-

---

[2] ϵ_o is a universal constant which can be taken as a potential energy unit. We can also write $M(z) = -2\mathrm{II}(z)$ if $\mathrm{II}(z)$ is expressed in electron volts.

late under the influence of this field. *No holes can be generated.*

2. *Assimilation of Superstandard Atmosphere with an Index Barrier*

This case is assimilated to that of a potential with a sudden discontinuity. Fig. 3 shows two examples of
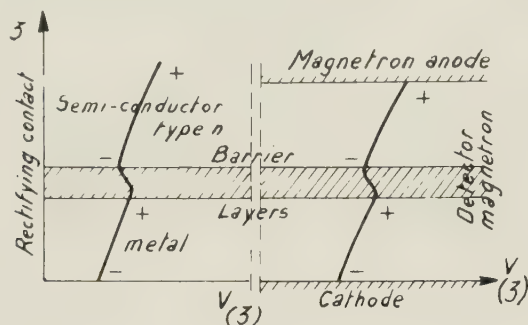

Fig. 3.

"potential barriers." The first is that of the potential barrier of an n type semiconductor—metal contact or a $p-n$ junction polarized in the direction of low impedance; the second is that of the potential barrier created by the narrow ring of space charge existing in a magnetron in the cut-off state (Gutton-Ortusi theory).[3] Fig. 3 is similar to the classical curve of the index profile with an inversion layer of Fig. 2.

It has then been stressed that, besides the Eckersley modes, certain Gamow modes are possible in that case for low values of μ_n. These modes are created by the passage into the inversion layer and are substantially independent of the boundary conditions.

There results, by assimilation, the following principle:

When, in a solid or in a vacuum, the electric field varies suddenly, some of the particles which cross the potential barrier *consist of holes.* These holes are complementary to electrons of low kinetic energy. *They are created in the barrier layer* representative of the potential barrier.

III. The Electronic Barrier Layers and Their Applications to Detector Magnetrons

We have examined, in Sections I and II, the properties of the regions of space with very high electron density. These properties are very close to the properties of solids. In particular, the zones of sudden variation of the electric field constitute, *even in vacuum, electronic barrier layers, generators of holes* and are analogous to the barrier layers which exist in $p-n$ junction in semiconductors.

These barrier layers exist in positive grid tubes in the neighborhood of the anode. We shall study their properties in magnetrons in the cutoff state or in traveling-wave tubes with crossed electric and magnetic fields.

Space Charge of the Magnetron in the Cutoff State [Fig. 4a)]

For its distribution, we shall take the hypothesis of H. Gutton and J. Ortusi.[4] The space charge is concentrated in a ring of very small thickness situated in the interelectrode space near the cathode [Fig. 4 (b)]. This ring of large charge constitutes an *electronic barrier layer* with all the properties of the surface barrier layers of solids, but with the difference that it possesses a resonance frequency ν_o given by the equation

$$2\pi\nu_o = \omega_o = \frac{eH}{m}. \tag{19}$$

This electronic barrier layer possesses the same rectifying properties as the barrier layers on the surface of separation of two semiconductors; *e.g.,* between two regions of germanium with positive and negative impurities ($p-n$ junction).
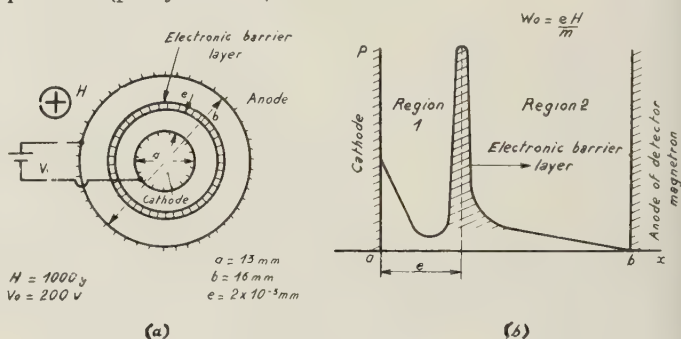

Fig. 4.

Assimilation with a p-n Junction—Mechanism of Rectification

1. Figs. 5 (a) and 5 (b) show this assimilation and explain the action of these two barrier layers, by the creation of complementary particles associated with evanescent waves [holes in Figs. 5 (a) and 5 (b)] and by their circulation during the positive phases of the ac voltage V.

2. Figs. 6 (a) and 6 (b) show the circuit equivalent to the two barrier layers. The ratio R_2/R_o can reach 10.[6] The $p-n$ junction, *fixed barrier layer*, possesses a limit upper frequency ω_e. An ac voltage is developed at its terminals if $\omega << \omega_e$. The electronic barrier layer, *mobile barrier layer*, possesses a resonance frequency. An ac voltage is developed at its terminals if $\omega \# \omega_o$.

3. Fig. 7 (a) shows a characteristic rectification curve of a $p-n$ junction. Curve 1 corresponds to the circulation of holes. Curve 3 corresponds to a circulation of free electrons by Zener effect. Fig. 7 (b) indicates in full lines, the rectification curve of a magnetron, expected by analogy with Fig. 7 (a), when $\omega \# \omega_o$; the origin O being situated on a point on the static curve of the magnetron (shown in dotted lines).

---

[3] This theory is explained in: "UHF modulation on waveguides" 1947 Radio Convention of the British Institution of Radio Engineers 20th May in Bornemouth, England; See also Onde Electrique, vol. 27, pp. 307-312, August 1947 and Fechner, Ann. Radioelect., vol. 28, pp. 83-105, April 1952.

[4] This hypothesis has been verified by the electron diffusion experiments of Reverdin, *Jour. Appl. Phys.,* vol. 22, p. 257–000; 1951. The curve of Fig. 4 (b) and the various resonance frequencies of the ring have been calculated and verified by Fechner, *Ann. Radioelect.,* vol. 29, pp. 199–220; July, 1952.

Fig. 5.

Fig. 6.

Fig. 7.

Fig. 8.

4. Fig. 8 (a) shows the variation of the rectified current as a function of $|V|$ in the case, rare in p-n junctions, where the slope of Curve 3 is greater than that of Curve 1 of Fig. 7 (a).

A curve is obtained which is rigorously linear for $|V| < V_z$ and a positive rectified signal I_{De} [according to the conventions of Fig. 5 (a)] followed by the inversion of the signal corresponding to the Zener effect.

Fig. 8 (b) shows the EXPERIMENTAL curve of the variation of the rectified current as a function of $|V|$ in a magnetron in the cutoff state ($I_{De} \equiv \Delta I_o$). This curve is *identical* to that of Fig. 8 (a). The region of positive linear rectification is obtained, in a full anode magnetron, for ac powers comprised between 10^{-12} and 10^{-9} watts. It corresponds to the circulation of holes. The region of negative rectification corresponds to the circulation of free electrons by a breakdown effect of the barrier layer analogous to the Zener effect.

Interpretation of the Experimental Curve of Fig. 8 (b).

1. The hypothesis of an electronic barrier layer with the properties of a solid allows of the complete prediction of all the characters of the experimental curve and of measuring the breakdown voltage V_z.

2. The hypothesis of a barrier layer on the cathode can explain neither the resonance frequency, nor the inversion of the rectified signal.

3. The hypothesis of the modulation of the ring can explain the phenomenon of resonance and negative rectification (the ring being very close to the cathode). It can explain neither positive rectification, nor its linearity for very weak signals.

4. Finally, in positive grid triodes, the same phenomena are observed experimentally in the neighborhood of the anode when its voltage is near zero. The double rectification, first positive, then negative, of Fig. 8 (b),

is observed. The resonance frequency of the electronic barrier layer here depends on the anode voltage.

Properties of Electronic Barrier Layers

Their properties are those of solids, to which they can be assimilated.

1. They present the phenomenon of *rectification* by barrier layer and *emission of holes*, as has just been seen.

2. They present the phenomenon of *dispersion of electronic waves* under the action of an electron bombardment with a diffusion zone replacing the Bragg rings. This phenomenon was demonstrated by Mr. Reverdin.[4]

3. They present the phenomenon of *luminescence*, under the action of fast electrons springing of cathode; i.e., they emit an ultra-violet radiation or even visible radiation, *even in an extreme vacuum*. This phenomenon is observed experimentally in cutoff magnetrons.

4. They present the phenomenon of *photoconductivity*; i.e., the hole current in the positive rectification zone can be greatly increased by the presence of an incident luminous flux. This phenomenon is observed experimentally. When the rectifying magnetron cathode is heated normally (1,000 degrees C.), an erratic and strong increase of the rectified current is observed which can be attributed to the presence of infra-red radiation. It is necessary to heat the magnetron cathode only to 450 degrees C. to bring out weak signals.

5. Fig. 9 shows how to utilize the photoconductivity of the electronic barrier layer in a traveling-wave tube with crossed electric and magnetic fields. The electronic barrier layer, shown dotted, comes into resonance by an uhf wave coming in from input U. The rectified hole current, measured between the anode and the sole-plate of the tube, varies when a flux of polarized light in the direction of the electron beam enters the interaction space through an aperture K. The current variation ΔI is proportional to the luminous flux Φ.
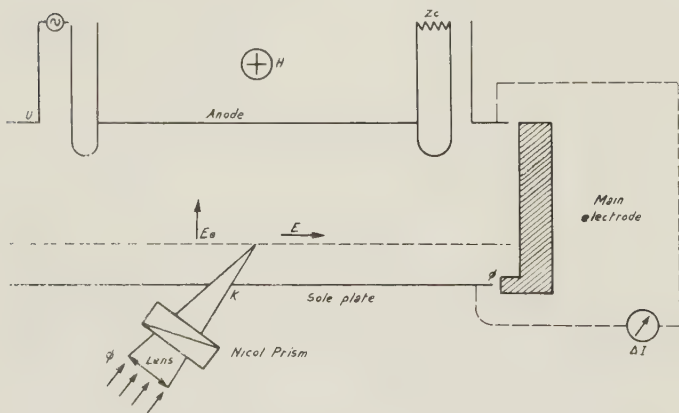

Fig. 9.

CONCLUSION

We have shown, both theoretically and experimentally, that zones with large space charge have properties equivalent to those of solids, with the advantage of being able to modify easily the velocities and the transit times of the positive or negative particles.

The electronic barrier layers have the same rectifying and photo-conducting properties as those of solids. Further, they present the phenomenon of resonance.

Use as a Rectifying Element

They have been used, at the Compagnie Générale de T.S.F. in Paris, in the design of a radar detecting system much more robust than a mixer crystal. To this end, the barrier layer of a magnetron in the cutoff state is employed as in Fig. 4. In order to increase the uhf signal on the cathode and to push back the value of the voltage V_z where the signal reverses [Fig. 8 (b)], various detecting tubes have been studied, obtained by modifying the cavities and interelectrode space of a multi-cavity magnetron.

Experiment has shown that the best magnetron detector obtained is one with 4 very open cavities (Fig. 10).
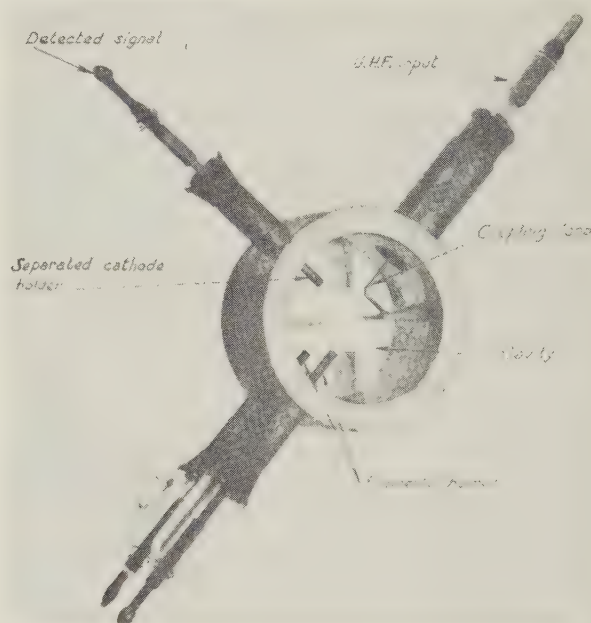

Fig. 10.

The anode and cathode diameters are very nearly equal (16 mm and 13 mm at 3,000 mc). For this type of tube, a particular orientation of the magnetic field pushes back considerably the limit V_z of linear rectification and justifies the use of this tube in a radar receiver. This tube has been associated with a video amplifier with a 1-mc bandwidth. It was observed that a uhf signal (at 3,000 mc) of 2 microseconds vanished in receiver noise only when its amplitude was 146 db below 1 watt. Unfortunately, the passband was very limited by the value of the magnetic field. It was of 2 to 3 mc with a fixed field, and of the order of 40 mc when the field was varied, while keeping the same tube. The rectified signal, examined with a standardized Hewlett-Packard generator is rigorously linear down to loss of the signal in noise.

Advantages of Detection by Means of a Magnetron. Detection by means of a magnetron presents three important advantages over conventional uhf detection by mixer crystal.

1. It does not require the use of a local oscillator or of an IF amplifier. A simple video amplifier is sufficient.

2. The noise factor, at 3,000 mc, is of the same order as that of crystals. Sensitivity, for a 1-microsecond pulse, is also the same. But, detection being rigorously linear whatever the frequency, sensitivity on millimetric waves becomes much higher than that of crystals.

3. During a radar transmission, magnetron detectors can withstand quite high powers. It is then unnecessary to provide a protection system, and a wideband TR with 30-db decoupling is sufficient.

Disadvantages of Magnetron Detection. On the other hand, magnetron detection has the following disadvantages:

1. The pass band is not very wide and, at 3,000 mc, sensitivity for narrow pulses (less than 0.1 μsec) is less than that of a conventional radar receiver. This disadvantage diminishes on millimetric waves.

2. The adjustment of the value and orientation of the magnetic field for increasing the inversion voltage of the signal V_z so as to receive all near echos, requires a very stable circuit.

3. The cathode temperature plays a very important role. With an ordinary cathode, it is observed that the weakest signal, observed above noise, is obtained for a temperature of 450 degrees C., at which the cathode does not glow. Below this temperature, the cathode has practically no emission. Above that temperature, infra-red radiation from the cathode causes considerable noise which originates in the photoconductivity of the electronic barrier layer. In these conditions, the much underheated cathode operates in a temperature region where the electron emission is unstable.

Detection by traveling wave tube with transverse magnetic field. These difficulties are much reduced when, instead of a magnetron, an equivalent traveling-wave tube is used (tube with crossed electric and magnetic fields as in Fig. 9). Noise introduced by the cathode can then be suppressed, and it is possible to produce a very large space charge enabling detection sensitivity to be considerably increased.

Use as a Photoconducting Element

Finally, electronic barrier layers enable the existence of photoelectric cells to be predicted in which no material element enters. No semiconductor is necessary to obtain the photoelectric effect. Although no experiment has been made, it would appear possible to obtain cells more suitable in operation than cells of the present type by avoiding the disadvantages inherent in the use of semiconductors (presence of impurities, rapid variations with temperature, etc. $\cdots$).

The study of electronic barrier layers in vacuum therefore appears to be a very important one, since they provide a link between solid-state physics and electrons in a vacuum.

Scattering at Oblique Incidence From Ionospheric Irregularities

D. K. BAILEY[†]

ABSTRACT

THE PURPOSE of this paper is to supply a review of the experimental investigations conducted by a research group at the National Bureau of Standards on the subject of oblique incidence ionospheric scatter at very high frequencies. Emphasis is placed on the results which contribute to further understanding of the behavior of the scattered signals and the physics of the ionosphere. The main topics discussed are long- and short-term characteristics of scattered signals. Included under the latter are the results of simultaneous observations at closely spaced receiving points. Next experimental techniques are discussed from which the heights of the ionospheric irregularities can be deduced and results of their use at different seasons and times of day are given. It is found that all the scatter which could be observed with the available techniques resulted from irregularities in the E region of the ionosphere, and most of it came from the lower parts of the E region. The following section deals with the dependence of the scattering on frequency; a range of frequencies from 27.775 mc to 107.8 mc is examined. The main part of the report concludes with the results of an experiment which gave information on the dependence of ionospheric scatter on the angle of scatter. A final section reviews the results of some simultaneous observations in Alaska dealing with the characteristics of ionospheric scattering in the auroral zone. This section emphasizes the different and new results from the Arctic.

[†] National Bureau of Standards, Washington, D. C.

Forward- and Back-Scattering From Certain Rough Surfaces

W. S. AMENT†

Summary—Heuristic relations are derived between the specular reflection coefficient, R, and the radar echoing power of rough surfaces in which induced current elements are constrained to radiate equal powers in the reflected ray's direction and back toward the radar. To the extent that currents in the surface and fields scattered by it are calculable through a self-consistent formulation, a simple Fresnel-zone computation of R shows that σ_o, the radar area per unit area of mean plane, is proportional to $|R^2|$ $\sin^2 \theta$, where θ is the angle incident rays make with the mean plane. It is plausibly assumed that large scatterers on the surface cast shadows with "beamwidth" proportional to radar wavelength λ; here the argument leads to $\sigma_o \propto (|R^2| \sin^2 \theta)/\lambda$. In two appendices the law $\sigma_o = 4 \sin^2 \theta$ is derived for a lossless surface obeying Lambert's law, and a known self-consistent "solution" of a rough surface problem is examined by three generally applicable criteria.

T HIS PAPER is concerned with a certain class of model rough surfaces, invented not because they resemble any surface of practical interest, but because their approximate effect on electromagnetic waves is calculable. In particular, for some of these models, one can readily obtain a relation between a specular reflection coefficient, R, and σ_o, the radar area per unit area of the surface.

We treat first a strictly two-dimensional case, because the figures are easy to draw and because a specific case has been worked out in some detail.[1] The model surface is built up of perfectly conducting half-planes parallel to the $x = 0$ plane of a rectangular coordinate system, all half-planes lying in $y < 0$ with edges in the "mean-plane" $y = 0$ (Fig. 1). We assume that the edges are
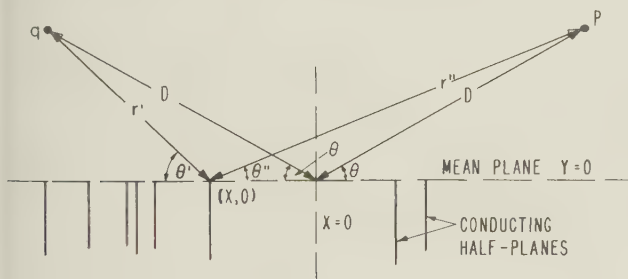

Fig. 1—A two-dimensional rough surface.

randomly and independently distributed along $y = 0$, so that N, the average number of edges per unit length of x axis, completely describes the "statistics" of the surface. Then we let a plane monochromatic wave $A \exp [ik(x \cos \theta - y \sin \theta)]$ fall on the surface from $y > 0$, and shall call this wave horizontally or vertically polarized according as its electric or magnetic vector is parallel to the z axis, i.e., to the edges.

This plane wave causes currents to be induced in the half-planes; the currents in turn re-radiate fields traveling in various directions. By symmetry the superposition of all such fields in $y > 0$ forms, on the average, a plane wave propagating according to $AR \exp [ik(x \cos \theta + y \sin \theta)]$, where R is again the specular reflection coefficient of the rough surface, evaluated as if the "mean-plane" $y = 0$ had formed a plane interface between two homogenous media.

Ascribing a power density CA^2 to the incident wave, we infer that the power density of the reflected wave is $CA^2 |R|^2$ so that energy conservation requires $R \leq 1$. The remaining power $CA^2(1 - |R|^2) \sin \theta$ is scattered randomly by an average unit length of the x axis. Part of this power passes downward between the half-planes, and some fraction of the remainder is "backscattered" in the direction of the source of the incident plane wave; the power may have flowed back and forth between several half-planes before finally returning into $y > 0$.

We are concerned with the relation between the power backscattered per unit length of surface and the specular reflection coefficient, R. Of all the randomly scattered energy, the backscattered is the easiest to calculate because, in the chosen geometry, the current sheet supported by each half-plane must radiate fields symmetrical with respect to the half-plane. The power radiated by the half-plane in the "backscatter" direction is then identical with that in the direction of specular reflection, and this is the key to establishing the relation.

To relate R with the backscattered power density, we first assume that average specularly and randomly reflected fields can be derived through a self-consistent formulation. That is, we assume that the field incident on a particular half-plane with edge at x_o may be regarded, to sufficient approximation, as the average field along (x_o, y) in the absence of the particular half-plane. Thus the currents in all half-planes are effectively assumed identical except for the phase factor $\exp (ikx \cos\theta)$, which contains the obvious dependance of the currents on the location of the half-plane's edge.

The second step in the argument is based on physical optics. We assume a line source at $q = (-D \cos \theta, D \sin \theta)$ and a receiving antenna at $p = (D \cos \theta, D \sin \theta)$, with $D = D(\theta_o)$ so large that for $\theta > \theta_o > 0$, any field reflected by the mean plane to p may be calculated by Fresnel zones to within preassigned accuracy. We further increase D, if necessary, so that all half-planes with edges in the first few zones can be regarded as illuminated by a plane wave with grazing angle θ. We then assume some amplitude factor, f, for the waves scattered per half-plane in the specular direction, and by symmetry, in the backscatter direction. Through Fresnel integrals we may then calculate R in terms of f,

[1] W. S. Ament, "Application of a Wiener-Hopf Technique to Certain Diffraction Problems," NRL Report No. 4334; May 10, 1954.

and the back-scattered power is given directly in terms of f. Eliminating f, we obtain the desired relation.

Quantitatively, we assume that the line source at q radiates a wave $\exp (ikr')(kr')^{-1/2}$ to a point $(x, 0)$ on the mean-plane at distance r' from q. This substantially plane wave is incident on $(x, 0)$ at grazing angle θ', and the point p is at distance r'' and at elevation angle θ'' from $(x, 0)$. Under the illumination from q, the half-plane with edge at $(x, 0)$ produces a field at p representable by

$$\phi(x) = f(\theta', \theta'') \exp [ik(r' + r'')]/(k^2 r' r'')^{1/2}, \quad (1)$$

where $f(\theta', \theta'')$ is a coefficient for scattering by a single half-plane in the θ'' direction when the plane illumination is from the θ' direction. We now have

$$R \exp (i2kD)/(2kD)^{1/2} = (\sum {}_i \phi(x_i))_{\text{average}}$$

$$= N \int_{-\infty}^{\infty} \phi(x)\, dx.$$

Evaluating the integral in the stationary-phase (Fresnel-zone) approximation, we obtain

$$R = N(2\pi i)^{1/2} f(\theta, \theta)/(k \sin \theta) \text{ or}$$

$$| f(\theta, \theta) |^2 = k^2 \sin^2 \theta | R |^2 /(2\pi N^2). \quad (2)$$

Now the field backscattered to q by a single plate with edge near $(0, 0)$ is approximately $\phi(0)$ times a phase factor which varies rapidly with the edge's distance from $(0, 0)$. Hence we may assume that the back-scattered power, per unit length of the mean plane near $(0, 0)$, and on the average over all arrangements of the half-planes, is simply $CN | \phi(0) |^2$. Defining (for this two-dimensional case) the "radar length per unit length of surface," as the dimensionless coefficient σ_L in

$$\frac{\sigma_L}{2\pi} = \frac{\text{(power backscattered, per radian, per unit length of surface)}}{\text{(power density of incident wave at the surface point)}}, \quad (3)$$

we obtain from (2)

$$\sigma_L = \frac{(2\pi D) | \phi(0) |^2}{(1/kD)} = 2\pi | f(0, 0) |^2 /k$$

$$= k \sin \theta | R^2 | /N. \quad (4)$$

Since θ_o and $D > D(\theta_o)$ were arbitrary, this relation holds, in general, to the extent that the self-consistent formulation is valid. (The validity is examined, for a special case, in Appendix B.)

Eq. (4) tells nothing about backscatter unless R is known; but for some extreme grazing incidence, determined by k and N, the surface should become a nearly perfect reflector, so that $| R | \approx 1$. Here

$$\sigma_L \approx 2\pi\theta^2/(N\lambda) \qquad (k = 2\pi/\lambda). \quad (5)$$

To a large degree, the very simple form assumed for the surface leads to the specific relation (4). Suppose that in a unit length in addition to the N half-planes per unit length already considered there were N' plates with edges in $y = 1$. These new "higher" half-planes would contribute relatively importantly to R, giving some scattering F as compared with the f of (1), (2),

which is now assumed unaltered. Then $R = Nf + N'F$, but $\sigma_L \propto Nf^2 + N'F^2$. Here F and f depend in different ways on θ, so that no clean-cut law like (4) is possible. But at extreme grazing incidence, only the currents in the "higher" half-planes may be significant, and $\sigma_L \propto \theta^2$ would still apply.

Now we turn to three-dimensional surfaces. The simplest analog of the vertical half-plane is now a thin wire lying in $y < 0$ parallel to the y axis; in the simplest thin-wire rough surface, the wire ends are randomly distributed in the mean plane, $y = 0$. The statistics of the surface are specified through n, the average number of wire ends per unit area of the mean plane. The "surface" now acts as a Faraday screen, so that the wires have little effect at horizontal polarization. For vertical polarization each (sufficiently thin) wire is the locus of a vertical current distribution having a radiation pattern axially symmetric about the wire. In particular, each thin wire scatters the same power in the "backscatter" as in the specular direction.

To obtain approximate solutions to the problem of diffraction by this surface, we may again use the self-consistent formulation to attack the plane-wave reflection problem. Each wire then has the same vertical current distribution, with phase again depending on the x co-ordinate of the wire's end through $\exp (ikx \cos \theta)$. Assuming a suitable *dipole* source at the point $q = (-D \cos \theta, D \sin \theta, 0)$ and a dipole receiving antenna at $p = (D \cos \theta, D \sin \theta, 0)$, we may again compute the average field at p by Fresnel zones, on the assumption that D is so large that illumination of the wires is substantially uniform throughout all significant zones.

The back-scattering properties of a three-dimensional rough surface are conventionally defined through the dimensionless parameter σ_o:

$$\frac{\sigma_o}{4\pi} = \frac{\text{[Power backscattered per steradian per unit area of the mean plane]}}{\text{[Power density in the incident wave at the unit area]}}. \quad (6)$$

For the present parallel-wire surface we use steps similar to those leading to (4), and obtain·

$$\sigma_o = \frac{| R^2 | k^2 \sin^2 \theta}{\pi n}. \quad (7)$$

At extreme grazing incidence ($\theta \approx 0$) we may expect that $| R | \approx 1$, so that

$$\sigma_o \approx \frac{4\pi\theta^2}{n\lambda^2}; \quad (8)$$

here the θ-dependence is in agreement with that of the ideal Lambert's-law surface discussed in Appendix A.

Thus far the results have depended on only two parameters: θ and N/k in (4) and (5), and θ and n/k^2 in (7) and (8). We now construct a three-dimensional parallel-wire surface from which a second k dependence of σ_o may be deduced. The thin wires are parallel to the y axis and extend from $y = 0$ to $y = -\infty$ as before, but this time they are arranged like the elements in cylinders,

as shown in Figs. 2(a) and 2(b). We assume that the adjacent wires in each "half-cylinder" are reasonably thin compared with their separation, so that each half-cylinder is built of a Faraday screen and is highly transparent to horizontally polarized waves. (The utility of this assumption will be made clear later.) Then the currents in each half-cylinder are constrained to flow parallel to the y axis, and the field re-radiated by each wire is axially symmetric about the wire, as before. For a nearly grazing, vertically polarized incident wave, the radiative interactions between the wires of a typical half-cylinder cause the half-cylinder as a whole to act approximately as a solid conductor. In particular, a single, isolated half-cylinder will cast a shadow of beam-width $\propto$ (wavelength/diameter), like that of Fig. 2(d).
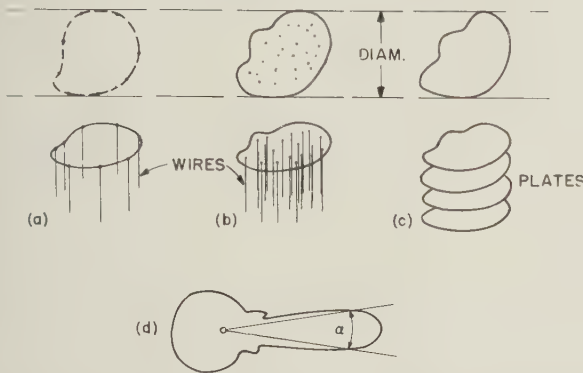
Fig. 2—Vertical structures producing plausibly related forward and backward scattering. In (d), the plausible polar scattering diagram for each of the structures includes a "shadow" of beam-width α proportional to (wavelength/diameter).

Again invoking the self-consistent hypothesis, we assume it approximately correct to regard the field incident on a particular half-cylinder of the surface as the average local field in the absence of the half-cylinder. It is plausible to regard this average field as approximating a bundle of plane waves propagating parallel to $z = 0$; and it is again plausible to regard resulting vertical currents in the wires of the half-cylinder as having complex amplitudes producing a combined radiation in the positive-x direction very like that of the forward radiation producing the shadow of the isolated half-cylinder. It is assumed even more plausibly that the radiation pattern in other than forward azimuths is independent of wavelength, as would be the case with large isolated, randomly corrugated or right-circular cylinders. We are then led to assume that the typical half-cylinder's scattering satisfies

Power scattered in specular direction, per steradian

Power backscattered, per steradian

$$\propto \text{Diam}/\lambda.$$

Assuming n' randomly located half-cylinders per unit area of mean plane and following the previous steps, we are led to

$$\sigma_o = \frac{\text{const} \cdot |R^2| \sin^2 \theta}{\lambda \cdot (\text{diameter}) \cdot n'}. \tag{9}$$

For grazing incidence, we again assume $|R| \simeq 1$ so that

$$\sigma_o \simeq \frac{\text{const} \cdot \theta^2}{\lambda \cdot \text{diameter} \cdot n'}. \tag{10}$$

We can similarly derive results formally identical with (9), (10) for horizontal polarization, by assuming that the Faraday-screen material composing the vertical half-cylinders consists of thin, separated conducting plates parallel to $y = 0$ as in Fig. 2(c).

In sum, the result (8) is plausible for vertical or horizontal polarization when the currents induced in the "surface" are constrained to flow vertically or horizontally, respectively. For half-cylinders having perfectly conducting vertical surfaces, currents can flow at some oblique angle with respect to the horizontal plane and the result (8) becomes, in the writer's opinion, much less plausible. The radiation from a vertical dipole is vertically polarized and axially symmetric about the dipole. The horizontally polarized radiation of a horizontal dipole is symmetric (at least powerwise) about the vertical through the dipole but varies otherwise with azimuth. (The higher symmetry for the "vertical" current elements makes (8) slightly more plausible, or at least more obvious, for vertical than for horizontal polarization, in the writer's mind.) The radiation pattern of an obliquely oriented dipole is unsymmetric in the general vertical plane intersecting the dipole, and in particular, a dipole parallel to the "specular" direction produces no radiation in that direction, but produces vertically polarized backscatter whenever $\theta > 0$. When the individual current elements are constrained to radiate symmetrically in the two directions of interest, then it seems plausible to combine the effects of the interacting elements in the half-cylinder into the radiation pattern of Fig. 2(b); but this plausibility appears lost when not even individual elements radiate symmetrically.

We therefore attempt no further detailed generalizations, but close with some remarks about the rough surface problem in general. The rough ocean is probably the surface of greatest practical interest. Here the wind orients the surface, as a back-scatterer, so that the downwind slopes near the wave crests are steeper than the upwind; the backscatter (as observed with airborne radars) is therefore considerably greater when the radars look into the wind than when they look downwind. (In addition, the microwave field between distant surface points is observed to be weaker than that predicted under the assumption of a smooth intervening ocean; hence the reflection coefficient of the rough ocean at nearly grazing incidence is of intrinsic practical importance.) The present randomly spaced parallel-plate surfaces may be "wind-oriented" by tilting all half-planes at some angle B with the y axis as in Fig. 3. Diffraction by this surface and by the analogous randomly distributed parallel wire "surface" can be formulated as Wiener-Hopf problems. The present symmetry arguments can be used to connect R with the randomly scattered power flux in the direction d of Fig. 3, but

other information (such as an approximate, analytic solution to the Wiener-Hopf problem) will be required for σ_L or σ_o.

The law $\sigma_o \propto \theta^2$ for $\theta << 1$, while that of the Lambert's-law surface, is in disagreement with other proposals that the limiting dependence should be as θ^3 or as θ^4. The present exponent 2 *may* arise because at $\theta \approx 0$, the incident rays strike the vertical structures nearly normally, for which reason the backscatter per half-plane, wire, or half-cylinder approaches some maximum. (It is interesting to note, from Fig. 3, that only for vertical half-planes and for $\theta = 0$ is the "symmetric" direction d that of geometric reflection by the half-planes.) The compensating effect is of course produced by the fact that the geometrically illuminated area of any half-plane diminishes as θ. The writer's guess is that the θ^2 law will obtain when the tilted half-plane problem is eventually tackled.
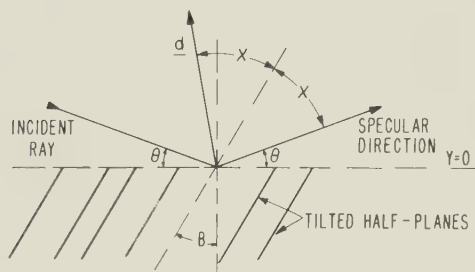
Fig. 3—A "wind-oriented" model rough surface.

Finally, we stretch the plausibility argument to deduce a $\sigma_o \propto \theta^2$ law for the general rough surface at extreme grazing incidence. When $0 < \theta < \theta_o \ll 1$ and the incident wave travels almost horizontally, it seems plausible to assume that $|R| \approx 1$ and that the illumination of each important scatterer on the rough surface varies, with θ, in *amplitude only* (except for the obvious phase-factor). Then the Fresnel-zone argument leads to $\sigma_o \propto \theta^2$. But the λ-dependence of the scattering may be quite complicated, so that a $1/\lambda$ variation of σ_o is much less plausible.

Appendix A

The Lambert's Law Surface as a Backscatterer

We devote some space to the derivation of the law $\sigma_o = 4 \sin^2 (\theta)$ for the lossless "perfectly rough," or Lambert's law, surface as the erroneous law $\sigma_o = 2 \sin (\theta)$ seems to permeate much of the classified literature on backscatter.

A surface satisfies Lambert's law of diffuse reflection if, under arbitrary illumination, a unit area of the surface scatters, per unit solid angle, a power proportional to the cosine of the (acute) angle between the direction of the scattering and the normal to the surface at the unit area. Thus a uniformly illuminated flat Lambert's-law surface of finite area produces, at an arbitrary point p in space, a scattered power flux proportional to the solid angle the area subtends at p, so long as the lit side of the surface can be seen from p.

Furthermore, the brightness (or apparent intensity) of the area is uniform throughout the subtended solid angle and independent of the position of p.

Now consider a unit area dA of a lossless Lambert's-law flat surface, illuminated at grazing angle θ with parallel light of unit intensity. The total power incident on dA is then $dA \sin (\theta)$. The power, scattered in a solid angle $d\phi \sin (\alpha) \, d\alpha$ at azimuth ϕ and zenith angle α, is 0 when $\pi/2 < \alpha \leq \pi$, and $I \cos (\alpha) \sin (\alpha) \, d\phi \, d\alpha$ when $0 \leq \alpha < \pi/2$. Here I is some proportionality constant to be determined by equating the total scattered power with the incident power, $dA \sin (\theta)$. It follows that $I = dA \sin (\theta)/\pi$, and the power E, per unit solid angle, scattered back along the incident beam is given by $E = dA \sin^2 (\theta)/\pi$. If dA had scattered isotropically, E would have been $dA \sin (\theta)/(4\pi)$ and σ_o would have been simply $\sin (\theta)$. If dA had scattered isotropically on the lit side only, with no scattering into the half-space shadowed by the plane tangent to dA, then E would have been $dA \sin (\theta)/(2\pi)$ and $\sigma_o = 2 \sin (\theta)$ would have held. Dividing the actual $E = dA \sin^2 (\theta)/\pi$ by $dA/(4\pi)$ we obtain $\sigma_o = 4 \sin^2 (\theta)$, as announced.

The erroneous law $\sigma_o = 2 \sin (\theta)$ appears based on the idea that a unit area dA of the ideally rough surface scatters uniformly over a hemisphere on the lit side of the plane tangent to dA. This assumption leads to a contradiction with either the first law of thermodynamics or with electromagnetic reciprocity, as follows: Let p and q be points equidistant from dA, with p on the normal to dA and q in a direction making the angle α with the normal. A dipole (parallel to dA) radiating unit power from p would cause some backscattered power P in the same dipole, and, under the assumption, would cause the same power in a suitably disposed dipole at q. By reciprocity, unit power radiation from the dipole at q would cause P at the dipole at p; by the assumption, the same power P is then backscattered into the dipole at q. (Hence, σ_o is independent of α, a contradiction already.) So that energy may be conserved, the assumed uniformly scattered power flux must be proportional to the power intercepted by dA. Hence the power at q due to p exceeds that at p due to q by a factor $1/\cos (\alpha)$, contradicting the conclusion from reciprocity.

One may argue that the reciprocity theorem has been misinterpreted by a factor $\cos (\alpha)$ in the foregoing. The assumption of uniform scattering still leads to a prediction contrary to the second law of thermodynamics. Suppose the illumination of dA is a hot wire of temperature T_o at p. A certain power flux P_o is then scattered from dA along rays diverging from dA. As seen from q, the apparent size of the source of these rays is $dA \cos (\alpha)$, so that a suitable lens system at q can focus a power CP_o (C a constant) into a hole, of area proportional to $dA \cos (\alpha)$, cut into a thermally insulated sphere which radiates only through the hole. If the interior of the sphere is at temperature T and radiates through the hole according to the black-body law, the radiated power is proportional to $T^4 dA \cos (\alpha)$,

i.e., to T^4 times the area of the hole. Under equilibrium conditions, this power is equal to the power CP_o entering the sphere. (The equilibrium T may be termed the apparent temperature of dA.) At equilibrium, T is proportional to the fourth root of sec (α) and can be made arbitrarily larger than T_o by taking α sufficiently near $\pi/2$. Thus power flows from a colder source to a hotter sink, in contradiction to the second law. (It is seen that if the scattered power flux is proportional to cos (α), as in Lambert's law, the "apparent temperature" of dA is independent of α.)

It may be objected that the radiation scattered by dA is at some finite mean wavelength, so that the image of dA must have some finite minimal size. But one may make the ray optics used above an arbitrarily good approximation by assuming, at the outset, that all linear dimensions are sufficiently large compared with the mean wavelength.

Finally, the law $\sigma_o \propto \sin(\theta)$ leads to the prediction that when a cube with ideally rough sides is illuminated solely by a flashlight held near the eye, the three visible sides should always appear equally bright. This significantly contradicts experience with cubes of any known substance.

APPENDIX B
DISCUSSION OF A PREVIOUS RESULT

For horizontally polarized plane waves incident on the parallel plate surface of Fig. 1, an approximate self-consistent solution valid for $\theta \ll 1$ and $Q = N/k \ll 1$ has been derived by a Wiener-Hopf technique.[1] The $f(\theta, \theta')$ and R of (1) and (2) were found to be of the form

$$f = (\text{const}) \left[\theta' + \sqrt{\theta^2 + 2iQ} \right] \qquad (11)$$

$$R = -[\theta - \sqrt{\theta^2 + 2iQ}]^2/(2iQ). \qquad (12)$$

From these and the present relation (4) was subsequently computed a $P(\theta, \theta')$, the power, per radian, scattered by a unit length of the surface in the θ' direction when a plane wave of unit power density is incident from the θ direction.

It was then possible to examine the solution for satisfaction of the criteria introduced in Appendix A: energy conservation, which requires that $P_r \equiv (1 - |R^2|)$ $\sin \theta - \int_0^{2\pi} P(\theta, \theta') \, d\theta' = 0$; reciprocity, which requires that $P(\theta, \theta') = P(\theta', \theta)$; and the second law of thermodynamics, which requires, loosely, that, for $\theta' > 0$, $P(\theta, \theta') \le C(\theta) \sin \theta'$, where C is some constant depending on θ. It was found that the residual power P_r vanished identically in an approximation compatible with $\theta \ll 1$; perhaps $P_r = 0$ is an identity for self-consistent formulations, and its calculation amounts to a check of the algebra. But the reciprocity criterion was not met except for the obvious $\theta = \theta'$, and, for $\theta \ne \theta'$, asymptotically for $Q \to 0$, where electromagnetic interactions between half-planes vanish. Finally, with $P(\theta, 0+) = P(\theta, 0) > 0$, the second law was not satisfied.

The formulation gave an approximate current distribution $I(\theta; y)$ in a typical half-plane. One can in principle use $I(\theta; y)$ and $I(\theta'; y)$ to obtain more accurate expressions for the randomly scattered power by an extension of Schwinger's variational principle to the present situation. Here reciprocity is satisfied automatically by the form of the variational expression, but whether P_r will be found to vanish is not clear.

NOTES ADDED IN PROOF

For a fixed plane incident wave, the current density at a fixed point in a fixed scatterer of the rough surface may be expressed at $J + J'$, where J is the current density averaged over all arrangements of the other scatterers, and J' has mean zero. In the self-consistent assumption, as used in the text and the reference, one ignores J' and calculates the specular reflection coefficient R and the randomly scattered flux as suitable averages, over all configurations, of reradiations from the currents J only. The R calculated in this manner is correct as the linearly superposed radiation fields of the J' vanish on the average. But in a calculation of the random flux, the omitted J' would appear quadratically, and a serious error is made. Mathematically, the error is akin to equating the average of the square of a random quantity with the square of its average. Physically, one could regard the J as the primary induced currents, then calculate the secondary currents J_2 induced (for a particular surface conformation) by the radiations from the J, then calculate the tertiary currents J_3 caused by radiations from J_2, etc.

This argument was developed shortly before the Michigan Symposium. In the talk actually given there, it was argued that $\sigma_L \propto \theta^4$ for sufficiently small θ. This conclusion is a consequence of Energy Conservation and Reciprocity as discussed in Appendix B, plus a reasonable assumption that $|R(\theta)| \simeq 1$, plus a plausibly argued assumption that $P(\theta, \theta') \propto f(\theta) f(\theta')$, both holding for small θ. The last assumption is that, as grazing incidence is approached, the non-specular power scattering pattern of the rough surface varies in intensity but not in shape; for the rather artificial surfaces of the present text, the equivalent assumption is that J varies in amplitude only for small θ. With $|R| \simeq 1$, the Fresnel zone arguments of the text then yield $J \propto \theta$, whence (from reciprocity and energy conservation) $\sigma_L \propto \theta^4$. We are trying to make these arguments semirigorous and to extend them to three-dimensional surfaces, where the statistical reciprocity relation is of a dyadic character.

The currents J_2, J_3, are neglected in the assumptions of the present text. This neglect is especially serious for random scattering in directions nearly tangent to the mean plane, where the radiation from a J is diffracted around other, randomly located scatters. Thus the errors, made in previous formulations[2,3] for $P(\theta, \theta')$ in terms of J only, appear "serious" only for small θ and/or θ'. The actual calculation of J' and its effects seems to be difficult.

[2] W. S. Ament, "Toward a theory of reflection by a rough surface," PROC., IRE, vol. 41, no. 1; January, 1953. (C. f., (13), (16))

[3] V. Twersky, "Multiple scattering of waves by planar random distributions of parallel cylinders and bosses," N. Y. U. Research Report No. EM-58 (C. f. "I" in (56))

Cerenkov and Undulator Radiation

H. MOTZ†

INTRODUCTION

During the last few years several authors have suggested applications of the radiation properties of electron beams to the problems of millimeter wave generation and some experiments have been carried out. Some investigators have succeeded in generating coherent waves in a range below one millimeter wave length. It is suggested to call these waves "Interwaves" or "Zwischenwellen" as they lie between the infra-red and microwaves. A separate name might be appropriate as the techniques for generation and detection are likely to be very different from microwave techniques.

The use of Cerenkov radiation has been suggested by Abele,[1] Danos,[2] and Linhart,[3] and tried,[4] and the radiation from fast electron beams has been investigated theoretically [Ginsburg,[5] Coleman,[6] Motz,[7] Combe and Feix,[8] and Landecker[9]] and experimentally.[10]

In this paper the theory underlying these devices will be examined from a unified point of view and the relationship between different approaches will emerge. It is hoped that the understanding of this rapidly increasing field will thereby be advanced. A related field, coherent light generation, will also be touched upon.

Cerenkov Radiation in a Homogeneous Medium

In 1934 Cerenkov[11] discovered a visible radiation emanating from pure liquids, including water, in the proximity of a radium source. His careful observations

† Oxford University, England.
[1] M. Abele, "L'effet Cerenkov en optique et dans le domaine des microondes," *Nuovo Cim. Suppl.* No. 3, to vol. 9, pp. 207–213; 1952.
[2] M. Danos, "Cerenkov radiation from extended electron beams," *Jour. Appl. Phys.*, vol. 26, pp. 2–7; 1955.
[3] J. G. Linhart, "Cerenkov radiation of electrons moving parallel to a dielectric boundary," *Jour. Appl. Phys.*, vol. 26, pp. 000–000; May, 1955.
[4] M. Danos, S. Geschwind, H. Lashinsky and A. Van Trier, "Cerenkov effect at microwave frequencies," *Phys. Rev.*, vol. 92, p. 828; 1953.
[5] V. L. Ginsburg, *Butl. Acad. Sci., U.S.S.R., Ser. Phys.*, vol. 9, pp. 165–182; 1947.
[6] P. D. Coleman, "Generation of Millimeter Waves," Ph.D. thesis, Physics Dept., M.I.T., 1951, and "Theory of Generation of Submillimeter waves—(1) Doppler effect," Elect. Engrg. Res. Lab. Rep., University of Illinois, Urbana, Ill.;
[7] H. Motz, "Applications of the radiation from fast electron beams," *Jour. Appl. Phys.*, vol. 22, pp. 527–535; 1951.
[8] R. Combe and M. Feix, "Mouvement d'un electron dans un ondulateur magnetique," *C. R. Acad. Sci.*, Paris, vol. 237, pp. 1318–1320; 1953, "Frequences et puissance des ondes rayonées dans un ondulateur magnetique," *C. R. Acad. Sci.*, vol. 237, pp. 1660–1662; 1953.
[9] K. Landecker, "Possibility of frequency multiplication and wave amplification by means of relativistic effect," *Phys. Rev.*, vol. 86, pp. 852–855; 1951.
[10] H. Motz, W. Thon, and R. N. Whitehurst, "Experiments on radiation by fast electron beams," *Jour. Appl. Phys.*, vol. 24, pp. 826–833; 1953.
[11] P. A. Cerenkov, *C. R. Acad. Sci.*, U.S.S.R., vol. 2, p. 451; 1934. For further bibliography see J. V. Jelly, "Cerenkov radiation," *Prog. Nuc. Phys.*, London, vol. 3, pp. 84–130; 1953.

convinced him that the phenomenon was not one of fluorescence. The proper explanation was found by Frank and Tamm[12] (1937).

It can be shown (e.g., Schiff, *Quantum Mechanics*, p. 261) that electrons in free space cannot radiate because energy and momentum cannot both be conserved in such a process. Radiation is possible when interaction with other charges involving momentum or energy transfer takes place. The other charges may be polarization charges in a dielectric or image charges in certain metal objects.

When an electron passes through a dielectric medium polarisation is momentarily produced and the medium reradiates transferred energy. One can see from the simple Huygens construction of Fig. 1 that a wave front will form in a direction θ such that the radiation travels from A to C. For a dielectric medium the particle velocity v may be larger than the wave velocity u and the latter is c/n when c is the speed of light in free space and n the refractive index.
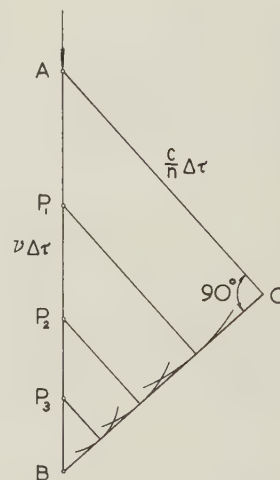

Fig. 1—Huygens construction for Cerenkov cone. Spherical waves starting at successive positions A, P_1, P_2, $\cdots$ along the electron path form a wavefront B–C.

According to the Huygens construction of Fig. 1,

$$c/n = v \cos \theta. \tag{1}$$

Eq. (1) defines an extremely thin conical shell in which the radiation is contained. The thickness of this shell is the smaller the larger the particle path is in the medium. The phenomenon is analogous to that of a shock wave emitted by a supersonic projectile or the bow waves of a ship. (In the latter case the wave velocity is that of surface waves.) Energy is radiated into the entire frequency band for which $n = (\epsilon)^{1/2} > 1$.

[12] I. Frank and Ig. Tamm, *C. R. Acad. Sci.*, U.S.S.R., vol. 14, p. 109; 1937.

This band stops at wave lengths longer than X-ray wave lengths where $n < 1$. The electron velocity v is of course not quite constant as the radiation energy must come from the particle. But the energy loss and speed change are small. Indeed if the phenomenon is to take place as stated the times it takes the electron to pass successive distances of one radiation wave length must not differ by more than a small fraction of the radiation period. Since the electron does not slow down appreciably, inertia effects are negligible and the radiated power therefore does not depend on the particle mass. It is the same for all particles of equal charge and speed (protons, mesons). The amount of energy radiated per unit of time is given by

$$W = e^2/c^2 \int_{v(\epsilon)^{1/2} > c} (1 - c^2/\epsilon v^2)\omega \, d\omega, \qquad (2)$$

where the integral is extended over all frequencies for which $v\{\epsilon(\omega)\}^{1/2} > c$. It is seen that this espression does not depend on the mass of the particle. Cerenkov radiation thus differs very markedly from radiation produced by acceleration or deceleration of charges where the intensity depends on $(e/m)^2$. We summarize the following points:

1. Cerenkov radiation is emitted in a homogeneous medium if the particle speed is equal to the wave velocity or exceeds it.

2. It is emitted in a thin conical shell of opening
$$\theta = \cos^{-1}(c/nv)$$

3. The energy radiated does not depend on the mass.

Radiation from Fast Accelerated Electrons in a Homogeneous Medium

We shall now contrast this with the properties of radiation from fast moving accelerated charges: $(v \cong u)$

1. This is contained within a narrow cone centered on the direction of movement of the charge with angular opening
$$\theta \cong \theta_o \qquad (3)$$
where $\theta_o = 1/\gamma$, $\gamma = 1/(1 - v^2/u^2)^{1/2}$, and u is the speed of radiation.

2. The wave lengths of radiation from electrons in sinusoidal motion of spatial period l_o range down to
$$\lambda_o \cong l_o/\gamma^2. \qquad (4)$$
The shortest wave lengths are propagated along the center of the cone and the radiation shades off towards higher wave lengths for larger angles.

3. For a charge rotating on a circle of radius R they range from high values down to
$$\lambda_o \cong R/\gamma^3. \qquad (5)$$

It is advantageous to get a simple understanding of these main features before one proceeds with the mathematical analysis.[13]

[13] L. I. Schiff, "Quatum effects in the radiation from accelerated relativistic electrons," *Amer. Jour. Phys.*, vol. 8, pp. 474–478; 1952.

Consider the radiation from an extended electron of speed v in accelerated motion in the direction $e - e'$ (Fig. 2) which is received at P. As long as the velocity component $v \cos \theta$ of the electron motion in the direction $e - P$ is near the radiation velocity u the radiation emitted later catches up with that emitted earlier and the intensity is increased. The time it takes a spherical wave S_1 of potential, centered at P, to sweep over the charge is a measure of the pile up of potential that will be observed when this wave finally collapses on P.
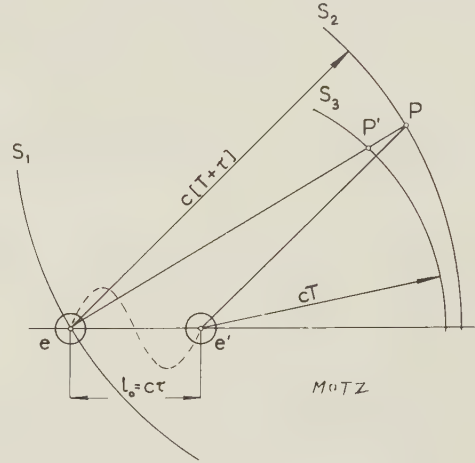

Fig. 2—Geometric construction for thickness of radiation pulse. The electron moves from e to e'.

This time is proportional to
$$u/(u - v \cos \theta) \cong 1/\{1 - v/u + \tfrac{1}{2}(v/u)\theta^2\}$$
$$= 2/(\theta_o^2 + \theta^2). \qquad (6)$$

This pile up is marked as long as $\theta = \theta_o$. Hence statement (1) follows. To illustrate the pile up further, two spheres S_2, S_3 have been drawn in Fig. 2. S_2 is centered at e and its radius is $c(T + \tau)$, S_3 is centered on e' and has radius cT. If the electron takes time τ to travel the distance l_o from e to e' the radiation emitted during this interval is piled up in the shell between S_2 and S_3.

It is easy to show that the thickness of the shell PP' is
$$u\tau(1 - v/u \cos \theta) = \tfrac{1}{2}u\tau(\theta_o^2 + \theta^2). \qquad (7)$$

Suppose now that the electron has carried out one period of a periodic motion between e and e' the radiation will contain prominent fourier components with period τ/γ^2 or wavelength
$$\lambda_o = u\tau/\gamma^2 = l_o/\gamma^2. \qquad (4)$$
This is statement (2).

Statement (3) follows from a consideration of Fig. 3. While the charge has rotated from e to e' a cone of radiation of angular opening $\theta_o = 1/\gamma$ sweeps past the observer during the time $R\theta_o/u$. One such pulse occurs during one revolution of the electron. This pulse has a thickness $= \tfrac{1}{2}u\tau\theta_o^2 = \tfrac{1}{2}R/\gamma^3$ and therefore high harmonics of the rotation frequency will be observed with wave lengths of this order.

In free space $u = c$. In a dielectric medium

$$u = c/n = c/(\epsilon)^{1/2}, \quad \gamma = 1/(1 - \epsilon v^2/c^2)^{1/2}. \quad (8)$$

Infinite frequencies seem possible when $\epsilon v^2/c^2 = 1$. It has however been mentioned before that dielectric properties change when high frequencies are reached and ϵ cannot exceed unity.
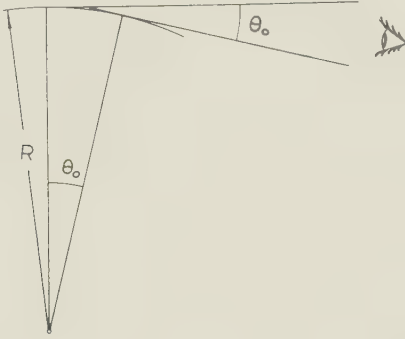

Fig. 3—Radiation cone emitted by an electron rotating on a circle of radius R.

APPLICATIONS

The idea underlying the method of generating millimeter and interwaves is the following: It becomes increasingly difficult at short wave lengths to use klystrons, magnetrons and traveling-wave tubes because the dimensions of these structures and the tolerances to be maintained in their fabrication become intolerably small. The undulator, however, does not require a structure of small dimensions for extracting high frequency power. The electron beam is passed through a succession of electric or magnetic fields of alternating polarity as in Fig. 4. The electron then follows a nearly
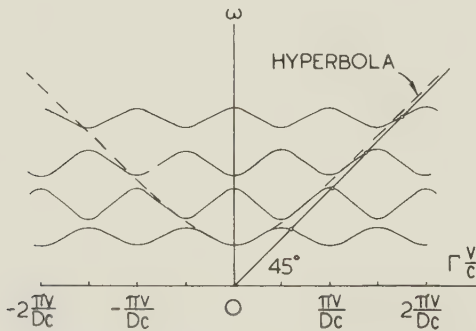
Fig. 4—Modified Brillouin diagram showing intersection of 45-degree line with phase shift curves drawn with a special scale factor. The intersections determine the radiated frequencies.

sinusoidal orbit in a transverse plane. The wave length is shorter than the period l_o of the magnet system by a factor $1/\gamma^2$. In the case of Cerenkov radiation, again, a small structure is not required if one can rely on the properties of a medium which will guarantee emission of short wave lengths as long as the dielectric properties exist in the millimeter wave region. There is, however, one difficulty inherent in this method. As the electrons

must travel outside the dielectric, they must tunnel through a hole cut through the dielectric or graze its surface. This feature tends to cut down the diameter of the electron beam which can efficiently contribute to the radiation.

MATHEMATICAL TREATMENT OF THE UNDULATOR AND CERENKOV RADIATION PROBLEM

In order to deal with the problem quantitatively we start with the wave equation for the transverse part of the vector potential A

$$\Delta A - (\epsilon/c^2)\partial^2/\partial t^2 A = -4\pi i/c. \quad (9)$$

The electron follows a sinusoidal motion and for high speed electrons the longitudinal speed variations are negligible so the electron orbit may be approximately given by

$$x = B \sin (2\pi/l_o)z, \quad z = vt, \quad (10)$$

and the current density i may be written in terms of Dirac δ-functions

$$i = ev\delta(z - vt)\delta(x - B \sin \omega_o t)\delta(y), \quad (11)$$

where $\omega_o = 2\pi v/l_o$.

Assume the expansion

$$A_{tr} = \sum_\lambda q_\lambda A_\lambda + q_\lambda^* A_\lambda^* \quad (13)$$

$$A_\lambda = (4\pi c^2)^{1/2} V^{-1/2} e_\lambda i^{(k\lambda r)}. \quad (14)$$

where the asterisk denotes conjugate complex quantities. The field is supposed to be enclosed in a volume V and the e_λ denote the direction of polarization. Negative and positive λ are included in the summation and we shall write $-k_\lambda = k_{-\lambda}$ and accordingly q_λ and $q_{-\lambda}$. Inserting (13), (14) (11) into (9), and multiplying by A_λ^* and integrating over space one finds

$$q_\lambda + q_{-\lambda}^* + \omega_\lambda^2(q_\lambda + q_{-\lambda}^*)$$
$$= e(e_\lambda \cdot v)/(\epsilon c V^{1/2}) \exp -ik_\lambda(vt \cos \theta +$$
$$\sin \theta \cos \phi B \sin \omega_o t)$$

$$\omega_\lambda = ck_\lambda/(\epsilon)^{1/2}, \quad (15)$$

where we have introduced polar angles θ, ϕ along the direction z. To obtain an equation for q_λ we note that the time dependence of q_λ has a sign opposite to that of $q_{-\lambda}^*$. Applying the operator $(i/\omega_\lambda) d/dt$ to (15) and adding the result to (15) we obtain

$$q_\lambda + \omega_\lambda^2 q_\lambda$$
$$= (1 + (i/\omega_\lambda) d/dt)e(e_\lambda \cdot v)/(\epsilon c V^{1/2})$$
$$\exp - ik_\lambda(vt \cos \theta + \sin \theta \cos \phi B \sin \omega_o t) \quad (16)$$

This equation is at once recognized as that of a forced oscillator. In the case of resonance the oscillator will continuously absorb energy. The energy of the oscillator represents the undulator radiation.

We expand

$$e^{\pm ik_\lambda \sin \theta \cos \phi B \sin \omega_o t}$$

$$= \sum_{p=-\alpha}^{\alpha} J_p(k_\lambda \sin \theta \cos \phi B)e^{\pm ip w_0}, \quad (17)$$

where $J_p(x)$ are the Bessel functions of order p.

Resonances are thus obtained when

$$\omega_\lambda = ck_\lambda/(\epsilon)^{1/2} = k_\lambda v \cos\theta + s\omega_o ; \qquad (18)$$

..e., when

$$\omega_\lambda = s\omega_o/(1 - [v(\epsilon)^{1/2}/c] \cos\theta) \qquad (18)$$

where s is an integer $(s = p,\ p + 1,\ p - 1)$.

It is interesting to note that the condition for resonance in the case $s = 0$ becomes $1 - v(\epsilon)^{1/2}/c\, \cos\theta = 0$. Thus this term represents the Cerenkov radiation. It is clear that the relation can only be satisfied if $\epsilon > 1$. This is the only term which would remain if the amplitude of the electron oscillation converged to zero.

Looking at (18) for the undulator frequencies one notes the characteristics of high speed electron radiation summarized before. When $\cos\theta \simeq 1$,

$$\omega \simeq 2\pi v/l_o(\theta_o^2 + \theta^2)$$
$$\lambda = 2l_o v/u\gamma^2 \simeq 2l_o/\gamma^2$$

The dependence on the mass comes in through the quantity B occurring in the argument of the Bessel function of (17). In the case of an array of magnets with maximum field strength H,

$$B = Hv[1 - (v/c)^2]^{1/2}/m_o\omega_o^2. \qquad (19)$$

It is seen that the q_λ for $s = 0$ depends on m. The Cerenkov term of the undulator equation thus violates one of the above criteria. Whether one confines the term to the case $B = 0$ is a question of terminology. The Bessel functions decrease for higher values of s, hence the higher harmonics are suppressed.

The power radiated can be calculated when the q_λ have been computed. The only force which does work on the particle is the electric field strength which is given by

$$E = -(1/c)A(z = vt). \qquad (20)$$

The work done per unit time must be equal to the power radiated and thus,

$$W = -(e/c)\Sigma q_\lambda\{v \cdot A_\lambda(vt)\} + q_\lambda^*\{v \cdot A_\lambda^*(vt)\}. \qquad (21)$$

In this manner one finds (2) for the case of the Cerenkov effect.[14] The case of the undulator in a homogeneous medium may be found in ref. (7). A detailed calculation will be demonstrated for the case of the radiation in a wave guide.

In order to derive a theoretical expression for the radiation of an undulating electron in a waveguide we start with the equation

$$\Delta\mathbf{\Pi} - (\epsilon/c^2)\mathbf{\Pi} = -4\pi r'\delta(r - r') \qquad (22)$$

for the Hertz vector π. In the waveguide we distinguish longitudinal and transverse co-ordinates. The co-ordinate vector is split up according to

$$r = r_{tr} + z. \qquad (23)$$

The electron coordinate is

$$r' = r_{tr}' + z' \qquad (24)$$

and $\delta(r - r')$ is written as $\delta(r_{tr} - r_{tr}')\delta(z - vt)$. For Π we now use the expansion

[14] A. Bohr, "Atomic Interaction in Penetration Phenomena," *Danske Vid. Selsk. Mat. Fys. Medd.*, vol. 24, pp. 1–51; 1948.

$$\mathbf{\Pi} = \int d\Gamma \sum_{n,m} e_v q_{n,m,}\ (\Gamma, t)\phi_{n,m,v}(r_{tr})e^{i\Gamma z} + C.C.$$

$$e_v = (e_x,\ e_y,\ e_z), \qquad (25)$$

are unit vectors in the co-ordinate directions, $C.C.$ stands for an expression which is the conjugate complex to that written down. The $\phi_{n,m}$ satisfy

$$\Delta_{tr}\phi_{n,m,v} + \alpha^2{}_{n,m}\phi_{n,m,v} = 0, \qquad (26)$$

where

$$\alpha^2{}_{n,m} = (m\pi/a)^2 + (m\pi/b)^2, \qquad (27)$$

in the case of a rectangular wave guide with sides a, b, $(a > b)$. They also satisfy the boundary condition and they are orthonormal according to

$$\int \phi_{n,m,r}(r)\phi_{n',m',v}(r)\ dr' = \delta_{n',m';n,m}. \qquad (28)$$

We shall subsequently often omit suffixes n, m. Substituting (25) into (22) we obtain

$$\int d\Gamma\ \big[-\ (\Gamma^2 + \alpha^2)q_{n,m,r}(\Gamma,\ t)$$
$$-\ q_{n,m,v}(\Gamma,\ t)\big]\phi_{n,m,r}(r_{tr})e^{i\Gamma z} + C.C.$$
$$= -4\pi e r'\delta(r_{tr} - r'{}_{tr})\ (\delta - vt).$$

We want to calculate $q_{n,m,v}(\Gamma,\ t)$. This is done by multiplying with $\phi_{n',m'}(r_{tr})e^{-i\Gamma'z}$ and integrating over the wave guide volume. Because of (28) and

$$\int e^{i(\Gamma \pm \Gamma')z}\ dz = 2\pi\delta(\Gamma \pm \Gamma') \qquad (29)$$

$$\int q_v(\Gamma,\ t)\delta(\Gamma \pm \Gamma')\ d\Gamma = q_v(\pm\Gamma',\ t) \qquad (30)$$

$$\int \phi_v(r_{tr})\delta(r_{tr} - r'{}_{tr})\ dr' = \phi_v(r'{}_{tr}), \qquad (31)$$

we obtain

$$(\Gamma^2 + \alpha^2)q_v(\Gamma,\ t) + (\epsilon/c^2)q_v(\Gamma,\ t) =$$
$$2e/c\{1 + (i/\omega_\Gamma)\ d/dt\}r_v'\phi(r'{}_{tr})e^{-i\Gamma'vt}, \qquad (32)$$

where we have assumed the time dependence

$$q_v(\Gamma,\ t) = q_v(\Gamma)e^{-iw\Gamma t}, \qquad (33)$$

and have eliminated $q(-\Gamma,\ t)$ according to the procedure of the previous section. We shall now proceed to solve (32) for the case of a rectangular guide. In this case,

$$\phi_{n,m,x} = \beta_m/(ab)^{1/2} \cos(n\pi x/a) \sin(m\pi y/b) \qquad (34)$$
$$\phi_{n,m,y} = \beta_m/(ab)^{1/2} \sin(n\pi x/a) \cos(m\pi y/b) \qquad (35)$$
$$\phi_{n,m,z} = \beta_m/(ab)^{1/2} \sin(n\pi x/a) \sin(m\pi y/b), \qquad (36)$$

where

$$\beta_m = \begin{cases} \sqrt{2} & \text{for}\quad m > 0 \\ 1 & \text{for}\quad m = 0 \end{cases} \qquad (37)$$

This follows because the boundary condition prescribes vanishing tangential components of electric field strength E. which is given by $E = \text{grad div } \Pi - (1/c^2)\Pi$. The z-component E_z must obviously depend on x and y as $\phi_{n,m,z}$. It may be verified that this is so when (34), (35), (36) are chosen.

The $\phi_{n,m,v}(r')$ contain trigonometric functions $\sin(n\pi x'/a) \cos(m\pi x'/a)$ which are expanded at $x' = a/2 - B \sin\omega_o t$. With the help of the expansion formulas

$$e^{\pm i(n\pi B/a)\sin w_0 t} = \sum_{p=-\infty}^{\infty} J_p(n\pi B/a)e^{\pm pw_0 t},$$

we obtain

$$\phi_{n,m,x}(\mathbf{r}') = \sum_p G_p e^{ip\omega_0 t} \tag{39}$$

$$\phi_{n,m,z}(\mathbf{r}') = \sum_p F_p e^{ip\omega_0 t}, \tag{40}$$

where

$$F_p = J_p(\pi nB/a)\begin{cases}(-)^n[\beta_m/2(ab)^{1/2}][1 + (-)^p] \\ \qquad\qquad \text{for } n \text{ odd} \\ (-)^{n/2+1}[\beta_m/2(ab)^{1/2}i][1 - (-)^p] \\ \qquad\qquad \text{for } n \text{ even} \end{cases} \tag{41}$$

$$G_p = J_p(\pi nB/a)\begin{cases}(-)^n[\beta_m/2(ab)^{1/2}i][1 - (-)^p] \\ \qquad\qquad \text{for } n \text{ odd} \\ (-)^{n/2+1}[\beta_m/2(ab)^{1/2}][1 + (-)^p] \\ \qquad\qquad \text{for } n \text{ even.}\end{cases} \tag{41a}$$

Eqs. (32) thus become

$$(\Gamma^2 + \alpha^2)q_{n,m,z} + (\epsilon/c^2)q_{n,m,z}$$
$$= (2e/c)[1 + (i/\omega_\Gamma)\,d/dt]\sum_p F_p vt e^{-i(\Gamma v - \omega_0 p)t} \tag{42}$$

$$(\Gamma^2 + \alpha^2)q_{n,m,x} + (\epsilon/c^2)q_{n,m,x}$$
$$= -(\omega_0 B)(e/c)[1 + (i/\omega_\Gamma)\,d/dt]\sum_p G_p\big[e^{-i\,[\,\Gamma v - (p+1)\omega_0\,]\,t}$$
$$+ e^{-i\,|\,\Gamma v - (p-1)\omega_0\,|\,t}\big] \tag{43}$$

We shall first solve (42). Defining a Fourier transform

$$q_z(\Gamma, \omega) = [1/(2\pi)^{1/2}]\int q_z(\Gamma, t)e^{i\omega\Gamma t}\,dt,$$

we find
since

$$\int e^{-i(\Gamma v - \omega_0 p - \omega)t}\,dt = 2\pi\delta(\Gamma v - \omega p - \omega) \tag{44}$$

and

$$\int te^{-i(\Gamma v - \omega_0 p - \omega)t}dt = 0, \tag{45}$$

$$q_z = 2e(2\pi)^{1/2}cv/e(\Lambda^2 - \omega_\Gamma^2)\Sigma F_p(i/\omega_\Gamma)\delta(\Gamma v - w_0 p - w_\Gamma); \tag{46}$$

where

$$\Lambda^2 = c^2(\Gamma^2 + \alpha^2)/\epsilon. \tag{47}$$

The solution (46) may be written

$$q_z = [e(2\pi)^{1/2}cv/\epsilon\Lambda](i/\omega_\Gamma)\sum_p F_p f(\omega_\Gamma), \tag{48}$$

where

$$f(\omega_\Gamma) = \{1/(\Lambda - \omega_\Gamma) + 1/(\Lambda + w_\Gamma)$$
$$+ i\pi[\delta(\Lambda - \omega_\Gamma) - \delta(\Lambda + \omega_\Gamma)]\}\delta(\Gamma v - \omega_0 p - \omega_\Gamma), \tag{48a}$$

where the sign of the last term is negative because the semicircle excluding $\omega_\Gamma = -\Lambda$ must be drawn above the real axis. This is so because we have assumed the time factor $e^{-i\omega_\Gamma t}$ so that a slightly attenuated (physical) wave must have a positive imaginary part on the negative axis.

The first two terms contribute the first term of the fourier inversion integral

$$q_z = 2evc\int d\omega \sum_p [F_p/\epsilon(\Lambda^2 - \omega^2)]e^{-i\omega t}(i/\omega)$$
$$\delta(\Gamma v - \omega_0 p - \omega)$$
$$- \pi evc \sum_p F_p/\epsilon\Lambda^2\begin{cases}e^{-i\Lambda t} \\ e^{i\Lambda t}\end{cases}\delta[\Lambda - (\Gamma v - \omega_0 p)]. \tag{49}$$

The first term has to be understood as the principle part of the integral, hence it vanishes in the expression for Π_z.

$$\Pi_z = \int \sum_{n,m} q_{n,m,z}\phi_{n,m},\ (\mathbf{r}_{tr})e^{i\Gamma z}d\Gamma + C.C.$$
$$= \pi evc\sum_{n,m}\sum_p \int (F_p/\epsilon(\Lambda)\Lambda^2)\delta[\Lambda - (\Gamma v - \omega_0 p)]$$
$$\phi(\mathbf{r}_{tr})e^{i(\Gamma z \pm i\Lambda t)}\,d\Gamma + C.C. \tag{50}$$

The δ-function requires

$$\Lambda = c[(\Gamma^2 + \alpha^2)/\epsilon]^{1/2} = \Gamma v - p\omega_0. \tag{51}$$

With the notation

$$1 - \epsilon v^2/c^2 = 1/\gamma^2,$$

we obtain for the undulator frequencies $\omega_{n,m}$ and propagation constants $\Gamma_{n,m}$

$$\omega_{n,m} = -\gamma^2\omega_0 p \pm \gamma^2 v(\epsilon\omega_0^2 p^2/c^2 - \alpha^2_{n,m}/\gamma^2)^{1/2} \tag{52}$$

$$\Gamma_{n,m,z} = -\gamma^2\epsilon\omega_0 pv/c^2 \pm \gamma^2(\epsilon\omega_0^2 p^2/c^2 - \alpha^2_{n,m}/\gamma^2)^{1/2}; \tag{53}$$

and for Π_z,

$$\Pi_z = -\pi evc\sum_{n,m}\sum_p (F_p/\epsilon(\omega_{n,m})\omega^2_{n,m})$$
$$e^{i(\Gamma_{n,m,z}z + \omega_{n,m,t})}\phi_{n,m}(\mathbf{r}_{tr}) + C.C. \tag{54}$$

The integer p can be positive, negative or zero. The last case deserves special attention. For $p = 0$,

$$\omega_{n,m} = i\alpha_{n,m}\gamma v = i\alpha_{n,m}v/\{1 - (\epsilon v^2/c^2)\}^{1/2} \tag{55}$$

$$\Gamma_{n,m} = i\alpha_{n,m}\gamma = i\alpha_{n,m}/(1 - \epsilon v^2/c^2)^{1/2}. \tag{56}$$

These frequencies are imaginary unless

$$\epsilon v^2/c^2 > 1; \tag{57}$$

then

$$\omega_{n,m} = \alpha_{n,m}v/(\epsilon v^2/c^2 - 1)^{1/2}. \tag{58}$$

If (57) is fulfilled we deal with Cerenkov radiation. It is seen that for each mode (n, m) the propagation velocity $u_{n,m}$;

$$u_{n,m} = \omega_{n,m}/\Gamma_{n,m} = v; \tag{59}$$

i.e., the wave velocity is equal and parallel to the particle velocity. It was mentioned in a preceding section that Cerenkov radiation in a homogeneous medium has a continuous spectrum. We now see that it has a discrete spectrum (58) in a waveguide. Cerenkov radiation, according to (54) and (41), is also emitted when there is no transverse electron motion; i.e., when

$$B = 0 \qquad J_0(0) = 1 \qquad J_p(0) = 0 \qquad p \neq 0.$$

The more general case of undulator radiation arises both for $\epsilon = 1$ and $\epsilon \neq 1$. Passing to the other values of p, we note that the solution (54) depends on z and t as

$$e^{i(\Gamma_{m,n}z \pm \omega_{m,n}t)}. \tag{60}$$

Solutions involving the positive sign correspond to incoming waves violating the radiation condition and

are therefore excluded. Eq. (60) can be written

$$\exp. i\gamma^2[-z\epsilon\omega_o pv/c^2 - \omega_o pt$$
$$\pm (\epsilon\omega_o^2 p^2/c^2 - \alpha^2/\gamma^2)^{1/2}(z - vt)]. \quad (61)$$

The root appearing as a factor of $z - vt$ becomes imaginary for high values of n and m—the waveguide cuts off.

One notes that the waves which are cut off in the frame of reference of the electron are attenuated in the laboratory frame of reference. This determines the choice of sign for the root in (52) and (53); for $z > vt$ the sign must be chosen positive, for $z < vt$ negative.

The calculation of the transverse component Π_x is rather similar to that of Π_z. The propagation constants and frequencies are given by

$$\Gamma_{n,m,q} = -\gamma\omega_o^2 v\epsilon(p \pm 1)/c^2$$
$$\pm \gamma^2[\epsilon\omega_o^2(p \pm 1)^2/c^2 - \alpha^2{}_{n,m}/\gamma^2]^{1/2} \quad (62)$$
$$\omega_{n,m,q} = -\gamma^2\omega_o(p \pm 1)$$
$$\pm \gamma^2[\epsilon\omega_o^2(p \pm 1)^2/c^2 - \alpha^2{}_{n,m}/\gamma^2]^{1/2}. \quad (63)$$

$q = 1$ corresponds to the plus sign

$q = 2$ corresponds to the minus sign.

It is interesting to note that the amplitude of the transverse component

$$\Pi_x = 2i\pi ce\omega_o B \sum_{n,m} \sum_{q=1,2} \sum_{p} (G_{p,q}/\epsilon\omega_{n,m,q})$$
$$\phi_{n,m}(r_{tr})e^{i(\Gamma_{n,mq}z - \omega_{n,mq}t)}, \quad (64)$$

has only one $\omega_{n,m,q}$ in the denominator, while the Π_z component is

$$\Pi_z = -\pi ec \sum_{n,m} \sum_{p} F_p/\epsilon(\omega_{n,m})^2$$
$$\phi_{n,m}(r_{tr})e^{i(\Gamma_{n,m}z - \omega_{n,m}t)}. \quad (65)$$

The power radiated can again be computed as the work done by the self force on the electron per unit of time. The electric field strength is

$$\mathbf{E} = \text{grad div.} \ddot{\boldsymbol{\Pi}} - (1/c^2)\ddot{\boldsymbol{\Pi}},$$

and the power is given by

$$P = e\{E_z(vt)v + E_x(vt)v_x\}; \quad (66)$$

where E_z and E_x are evaluated at the electron

$$(z = vt, x = (a/2) - B \sin \omega_o t).$$

The formulas in this section have been derived for an infinitely long undulator. In the case of a finite undulator of length L, (44) should be replaced by

$$\{\sin (\Gamma v - \omega_o p - \omega)L/2u\}/(\Gamma v - \omega_o p - \omega), \quad (44a)$$

and the last δ-function in (48a) is consequently replaced by (44a) divided by 2π. The spectrum is thus continuous. The discussion of this case will be given in a later publication.

RADIATION IN PERIODIC STRUCTURES

We have shown that Cerenkov radiation is emitted in a wave guide only when $\epsilon > 1$. This condition is no longer necessary if the waveguide is iris-loaded, or corrugated. Such guides, well known as slow wave structures, are used in traveling-wave tubes and linear accelerators. To show that Cerenkov radiation is emitted when $\epsilon = 1$ in periodic structures we start again

with (22) for Π. According to a theorem by Floquet (extended by F. Bloch) solutions can be written in the form

$$\Pi_z(r, T) = e^{i\Gamma_o z}f(z)\phi(r_{tr}) + C.C., \quad (67)$$

where $f(z)$ is periodic in z with the period D of the structure

$$-\pi/D < \Gamma_o < \pi/D: \quad (68)$$

for $f(z)$ we can write

$$f(z) = \sum_{m=o}^{\infty} g^{(m)}e^{(2\pi im/D)z} \quad (69)$$

$$g^{(m)} = i/D \int_o^D f(z)e^{-(2\pi im/D)z}. \quad (70)$$

To make the problem more specific consider a cylindrical guide of radius $r = a$ with slots of width d cut to a depth of radius $r = b$. The distance of successive slots is D. A simplified treatment, due to Walkinshaw,[15] of wave propagation in such a guide is found in H. Motz, (1951[16]). The solution for $r < a$ may be written

$$\Pi_z = \Sigma g^m e^{i[\Gamma z - \omega t]}J_o(\alpha_m r) + C.C. \quad (71)$$
$$\alpha_m^2 = \omega^2/c^2 - \Gamma^2 \quad (72)$$
$$\Gamma = \Gamma_o + 2\pi m/D. \quad (73)$$

The field in the slots for $r > a$ will be of the form

$$\Pi_z = \sum_p P_p[J_o(\alpha_p r) - N_o(\alpha_p r) J_o(\alpha_p b)/N_o(\alpha_p b)]$$
$$\sin \{2\pi p z_n/d\}e^{iwt} + C.C. \quad (74)$$
$$\alpha^2 p = \omega^2/c^2 - p^2\pi^2/d^2, \quad (75)$$

where z_n is the axial co-ordinate in the nth slot. The boundary condition at $z = b$ and at the slot walls is already satisfied by (74). Now the fields have to be matched at the slot mouths. This can be done by assuming a quasistatic field at the nth slot mouth:

$$E_z = C^{\pm i\lambda_o nD}\{(1 - (2z_n d)^2\}^{1/2}. \quad (76)$$

The electric field valid for $r < a$ is equated to (76) and the $g^{(m)}$ are determined in terms of C. The $g^{(m)}$ are inserted into the expression for the magnetic field H_r. Then the electric field valid for $a < r < b$ is equated to (76) and the constants P are determined in terms of C. Now the magnetic field valid for $a < r < b$ is determined in terms of C and equated to an averaged field valid for $r < a$. With certain further approximations this leads to an equation

$$F_1(\alpha_{p=o}a)/F_0(\alpha_{p=o}a)$$
$$= d/D \sum_{-\infty}^{\infty} (\alpha_{p=o}/\alpha_m)\{J_1(\alpha_m a)/J_o(\alpha_m a)\}$$
$$J_o(\gamma_m d/2)\{(\sin \gamma_m d/2)/(\gamma_m d/2)\}, \quad (77)$$

where

$$F_1(\alpha r) = N_o(\alpha b)J_1(\alpha r) - N_1(\alpha r)J_o(\alpha b)\cdots \quad (78)$$
$$F_o(\alpha r) = J_o(\alpha r)N_o(\alpha b) - J_o(\alpha b)N_o(\alpha r)\cdots. \quad (79)$$

It is clear from the preceding sections that the forcing

[15] W. Walkinshaw, "Theoretical design of linear accelerator for electrons," *Proc. Phys. Soc.* (London), vol. 61, p. 246; 1948.
[16] H. Motz, "Electromagnetic Problems of Microwave Theory," Methuen, London, p. 122; 1951.

frequencies are Γv. Putting this into (72), (75), we obtain

$$\alpha^2{}_m = \Gamma^2(v^2/c^2 - 1) \qquad (80)$$

$$\alpha^2{}_p = \Gamma^2 v^2/c^2 - p^2\pi^2/d^2, \qquad (81)$$

where Γ is given by (73). The transcendental equation (77) together with (73) determines Γ_o. The roots $\Gamma_o{}^i$ determine the Cerenkov frequencies

$$\omega^c{}_{i,m} = (\Gamma_o{}^i + 2\pi m/D)v. \qquad (82)$$

Eq. (77) is only approximate, the full matching condition is more involved.

In general one can represent the relation of phase shift versus frequency for a periodic structure in the familiar Brillouin diagram. If we draw a modified diagram with $\Gamma v/c$ as abscissa and ω as the ordinate, the emitted frequencies are given by intersections of a 45-degree line with the phase shift curves, Fig. 4.

RADIATION IN A GUIDE PARTIALLY FILLED WITH DIELECTRIC

Cerenkov radiation may also be emitted in a guide partially filled with dielectric so as to allow for the free passage of an electron beam. Again, to make the problem more specific, let a cylindrical wave guide of radius b be evacuated for $0 < r < a$ and lined with a dielectric for $a < r < b$. In the space $0 < r < a$ the solution will be

$$\Pi^2 = -jAe^{i(\Gamma z-\omega t)}J_o(\alpha r) + C.C. \qquad (83)$$

$$\alpha^2 = \omega^2/c^2 - \Gamma^2. \qquad (84)$$

For $a < r < b$ the field will be given by

$$\pi_z = -jBe^{i(\Gamma z-\omega t)}[J_o(\alpha_1 r) - N(\alpha_1 r)\,J_o(\alpha_1 b)/N_o(\alpha_1 b)] \quad (85)$$

$$\alpha_1{}^2 = (\epsilon\omega^2/c^2) - \Gamma^2. \qquad (86)$$

Along the boundary at $r = a$ the circumferential magnetic fields H_θ and the axial electric fields E_z in both regions must be equated. These conditions lead to the following equation (This equation is given by Abele[1] without derivation).

$$J_o(\alpha a)/J_1(\alpha a) = (\alpha_1/\epsilon\alpha)F_o(\alpha_1 a)/F_1(\alpha_1 a), \quad (87)$$

where F_o, F_1 are given by (78), (79). It is again clear that the forcing frequency is $\omega = \Gamma v$, so that

$$\alpha^2 = \Gamma^2(v^2/c^2 - 1) \qquad (88)$$

$$\alpha_1{}^2 = \Gamma^2(v^2\epsilon/c^2 - 1). \qquad (89)$$

Substituting (88), (89) into (87) we obtain a transcendental equation for Γ.

Since $\alpha = \Gamma\{(v/c)^2 - 1\}$ is imaginary, we write instead of (87),

$$I_o(x)/I_1(x)$$
$$= (\epsilon v^2/c^2 - 1)^{1/2}/\epsilon(1 - v^2/c^2)F_o(x\gamma_1/\gamma i)/F_1(x\gamma_1/\gamma i), (90)$$

with the notation

$$1/\gamma = (1 - \epsilon v^2/c^2)^{1/2} \qquad (8)$$

$$1/\gamma_1 = (1 - v^2/c^2)^{1/2}. \qquad (8a)$$

In terms of a solution ξ_i of (90), the Cerenkov frequency spectrum is given by

$$\omega_i{}^c = \xi_i v/a(1 - v^2/c^2)^{1/2}. \qquad (91)$$

It should be noted that the arguments of F_o, F_1 are real, i appears because of our previous notation. It is seen immediately from (90) that there is no solution when $\epsilon < 1$.

SOME OPTICAL EXPERIMENTS

The undulator theory was first confirmed by an optical experiment.[10] A magnetic undulator was used. It consists of an assembly of magnets arranged along a main axis so that their fields are perpendicular to the axis and alternate periodically with axial distance
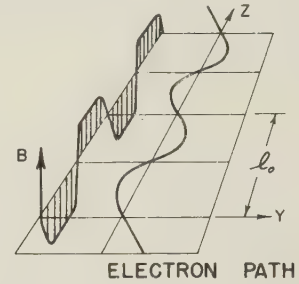

Fig. 5—Magnetic flux curve and electron path of a magnetic undulator. (*Previously published in Jour. Appl. Phys.*)

(Fig. 5). Provided the field contains only odd harmonics in z, an electron proceeding along this axis will describe a periodic trajectory if injected in a plane perpendicular to the field ($y - z$ plane) and at a certain angle with the axis. The undulator partially assembled is shown in
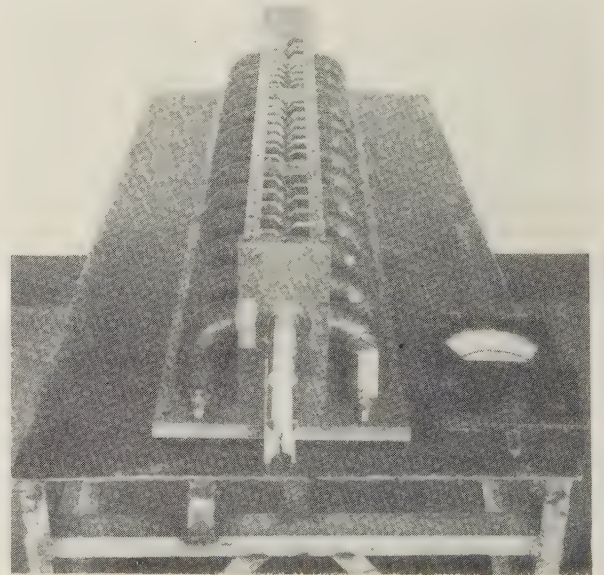

Fig. 6—Undulator structure, partially assembled. (*Previously published in Jour. Appl. Phys.*)

Fig. 6. An electron beam of 100 mev energy was available from the Stanford linear accelerator. Eq. (18) predicts that such electrons will generate waves in the visible range when passing through the undulator. This was observed and the polarization and the color of the light was found as expected.

An interesting counterpart of this experiment was devised by Smith and Purcell.[17] Light is generated by an electron beam of 300 kv passing close to the surface of a metal grating. We have given the theory of Cerenkov radiation in the case of the corrugated cylinder. In the case of a corrugated plane the theory leads to the undulator formula

$$\lambda = dc\{1 - (v/c)\cos\theta\}/v$$

for the fundamental wavelength.

The spacing of the grating, d, was 1.67 microns in this experiment. Smith and Purcell were able to record the spectrum of the emitted light and they proved that the dependence on beam energy and the polarization agreed with theoretical expectations. The beam carries out a slight undulating motion but the light is mainly due to the oscillation of induced surface charges and might thus be classified as a special kind of Cerenkov radiation.

Millimeter Wave and Interwave Experiments

It seemed interesting to pursue the subject of millimeter wave generation experimentally. As is well known, there are no sources of coherent radiation available in the millimeter- and sub-millimeter-wave range. Some workers have produced incoherent radiation by the so-called mass radiators. One thinks here notably of Glagolewa-Arkadiewa,[18] and also some more recent work by Cooley and Rohrbaugh[19] and others. This kind of radiation is incoherent in the sense that it could not be received by heterodyne detection, or else in the sense that radiation produced by one mass oscillator would not produce interference phenomena with the radiation produced by another mass oscillator. Undulator radiation is different in that respect. Radiation is coherent. The radiation from one undulator would interfere with the radiation by another undulator. Many uses of coherent radiation in the millimeter and submillimeter band come to mind. Naturally, it could have application in the communications field. It would have applications in radar. It would certainly be very interesting to study the properties of matter in that frequency range Availability of such radiation would open up an entirely new field of microwave spectroscopy.

The first experiments done at Stanford were designed to prove the principle in general. A magnetic structure was fabricated. Theory shows that the radiated intensity increases with the square of the magnetic field. Hence, it was desirable to make an undulator field of high intensity. This requirement was met by using magnetron magnets and putting an array of steel teeth between the magnet poles, so that an alternating flux was obtained.

The output of radiation is very much enhanced if the bunches of radiating electrons are very short. This is to say, if the bunches are short compared to the wavelengths received in the laboratory system, the intensity of the radiation emitted is proportional to the square of, rather than linearly proportional to, the current. Thus a factor of 10^3 to 10^4 or even more can be attained by which the radiation emitted by highly bunched electrons is more intense than the radiation emitted by an unbunched electron beam. Such bunching action seems possible in the millimeter- and to a certain extent in the sub-millimeter-wave range. Experiments on wave generation in that range were undertaken by means of a special buncher accelerator constructed for use with the Mark III (220-foot) linear accelerator at Stanford University. This buncher or linear accelerator is capable of delivering 3–5 mev with good bunching action. It is a structure in which the phase velocity varied from 0.5 times to virtually the velocity of light within about 3 or 4 feet. At the time when these early experiments were done, no spectrometer was available which could distinguish between waves of different frequency in the millimeter- and sub-millimeter-wave range. The detection was effected by crystal detectors which were manufactured in the laboratory. They were crystal detectors with especially fine tungsten "whiskers." These detectors could be evacuated, so that the radiation did not pass through any windows through which transmission was uncertain. Power in the millimeter-wave range was obtained and it was estimated that it was about 1 w or more. An instrument incorporating a thermistor was also used to estimate the intensity. A block diagram of the set up used in these experiments is
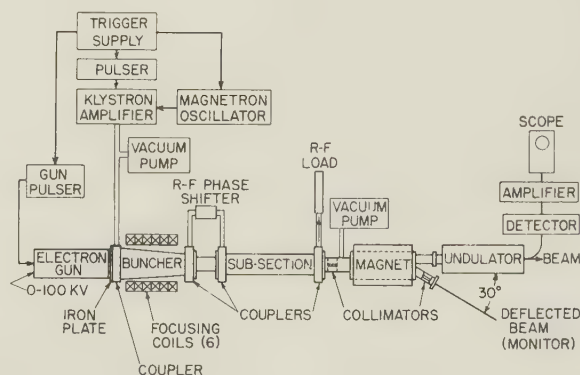

Fig. 7—Block diagram of undulator set-up. (*Previously published in Jour. Appl. Phys.*)

shown in Fig. 7. These experiments seemed encouraging and as the buncher used was needed by the Mark III accelerator, it was decided to build an accelerator which would be set aside for the purposes of the undulator experiments specifically.

Several conclusions were derived from the early experiments. First of all, it was concluded that further improvement of the bunching action would be desirable.

---

[17] S. J. Smith and E. M. Purcell, "Visible light from localized surface charges moving across a grating," *Phys. Rev.*, vol. 92, p. 1069; 1953.

[18] A Glagolewa-Arkadiewa, "Short electromagnetic waves of wave length up to 82 microns," *Nature*, vol. 113, p. 640; 1924.

[19] J. P. Cooley and J. H. Rohrbaugh, "The production of extremely short electromagnetic waves," *Phys. Rev.*, vol. 67, pp. 296–297; 1945.

It was also concluded that an electromagnet would be preferable to an arrangement of permanent magnets. The fact that one cannot switch off a permanent magnet caused a great deal of inconvenience in the early experiments. Moreover, it is not easy to trace the beam through the structure. It was ascertained that the beam did not penetrate the whole length of our first undulator, which was 3 feet long. Therefore, the experiments were first done with another, short model of an electromagnetic undulator, and a long strong focusing undulator was designed. A special prebuncher was constructed. This prebuncher consisted of two low-voltage cavities and one high-voltage cavity for accelerating the beam. The prebuncher would confine the bunches to within, let us say, 30 degrees; the accelerator would then take over and bunch the electrons further to within a few degrees. Detailed theory of such a prebuncher has been described by Thon.[20] The experiments done on millimeter-wave and visible-light generation have likewise been described.[7]

The more recent group of experiments was performed with an accelerator originally intended for a radiological application. The device was supposed to give an output of 3 mev with an input of 0.5 Mw, to be obtained from a magnetron. It proved, unfortunately, that the pipe was faulty, and therefore had attenuation somewhere along the structure. As a consequence the maximum energy which could be obtained with a power input of 0.5 Mw was about 0.7 mev. It was decided to supply the accelerator with 1 Mw from a Type 272 S-band klystron. With this level of power input, the output was somewhat higher than 2 mev, but 3 mev could not be reached. The simple theoretical analysis given above would show that 2.5 mev would be required if any radiation was to be produced in the undulator which had been constructed. The period of this undulator is 40 mm. At an energy of 2 mev this period would appear to an observer in the electron's frame of reference as an 8-mm. period. The electrons moved in a wave guide passing through the undulator. The cutoff wavelength of the waveguide was 7.4 mm. The wave guide should be effective for shielding of all radiation above 7.4 mm. in the electron frame of reference. It was hoped, however, that there would be harmonics in the flux curve of the undulator and that output might nevertheless be achieved. With electron energies of 2 mev and below, high-frequency radiation was indeed received. For the frequency analysis, an echelettetype spectrometer was available, constructed by R. N. Whitehurst according to designs contained in a paper by Strong.[21] This echelette spectrometer, Fig. 8, allowed wavelength deter-

minations to within a fraction of 1 per cent. There are four ruled aluminum plates. The first plate covers the range from 0.5 to 1 mm, the second one from 1 to 2 mm, the third from 2 to 4 mm, and the fourth from 4 to 8 mm. Some experiments on harmonic generation from K-band generators had been done previously with this spectrometer, and it was found satisfactory. The radiation received from an experimental 4.5- to 6-mm backward-wave oscillator was also measured on this spectrometer. The device seemed to be free from parasitic lines
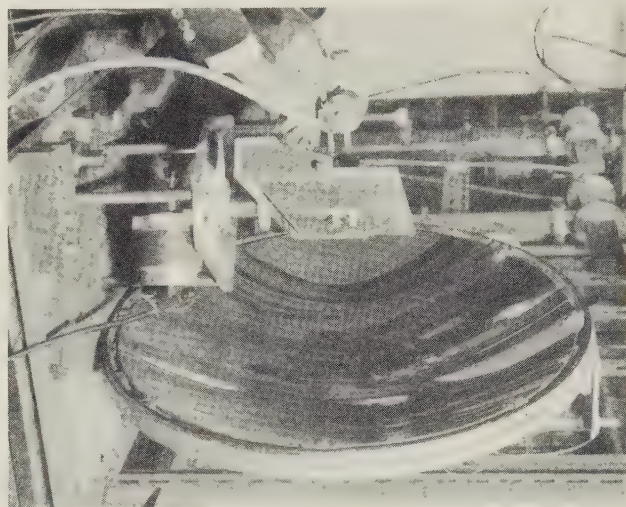

Fig. 8—Echelette spectrometer.

in the region tested. The finest plate, intended for the range from 0.5 to 1 mm, had of course not been tested yet.

On this plate one can also measure wavelengths smaller than 0.5 mm, but certain advantages inherent in the spectrometer design are lost in that range. The spectrometer construction is such that if a given order (say, the nth) of the interference is received, the $n-1$ and $n+1$ orders are inhibited, because they fall on the minimum of the Fraunhofer diffraction curve. (This phenomenon would be probably not observed in waves below the design range in wavelength.) The only result of this effect, however, should be that power reflected by the echelette plate is not as high as in the design range. In order to use the spectrometer, it was first necessary to make the radiation pass out of the evacuated waveguide into an air-filled wave guide. This was achieved by means of a 0.002-inch Milar window. Although the absorption by this window is unknown, it does not seem to weaken the signal to an appreciable extent. The spectral transmission of the window, however, should be tested. The spectrum measured by means of the echelette spectrometer is of the following type. On all of the plates a number of very sharp lines appears. On the plate intended for 0.5 to 1 mm numerous lines were observed even below 0.5 mm. The spectrum

[20] W. Thon, "Design, Construction, and Operation of a Magnetic Undulator," Stanford University Ph.D. thesis (in preparation); esp. ch. 6.

[21] J. Strong, "Resolving-power limitations of grating and prism spectrometer," *Jour. Opt. Soc. Amer.*, vol. 39, pp. 320–323; 1949.

ranges from 0.12 mm to the longest wavelengths (6.55 mm) received on the plate intended for the range 4 to 8 mm. There is at present no good explanation of this line. There are also regions of a continuous spectrum.

As an example of a measurement run we give the data in Table I:

TABLE I
Energy 1.64 mev finest grating, with magnet

Wavelength in mm	0.147	0.211	0.270	0.310	0.340
Approximate rel. intensity	2	3/2	1	1	3/2
Wavelength in mm	0.372	0.397	0.418	0.447	0.475
Approximate rel. intensity	1¾	1¼	1	1	½

Energy 1.5 mev coarsest grating

With undulator magnet	With undulator magnet
0.970	0.883
1.65	
1.85	
2.31	2.42
2.80	
3.15	
3.40	3.56
3.91	
4.95	4.88
5.85	
6.57	6.90
7.06	7.63
7.71	8.09
	8.49
8.85	8.72

The most intense lines on the plate for the shortest waves seem to be near the position in which the plate is parallel to the mirror. The explanation may be the short coherence length of the radiation made by the short undulator. For very small inclinations the rulings on the plate may all interfere together, but for larger inclinations only groups of echelette faces near to each other contribute. This circumstance may explain why no lines were found for the larger inclination angles between the echelette plate and the plane of the mirror.

Estimates of the measured power of the radiation are based on the same argument as used earlier by Motz.[7] Total power of high-frequency radiation is again estimated to be in excess of 1 w by comparison with threshold signals equal to noise which have been accurately measured by the Columbia University group. Coherent radiation in a wavelength range from, let us say, 0.1 to 1 mm has not been obtained before. It is hoped that a better understanding of the mechanism of oscillation generation will be obtained when further experiments are available. Some tests have been carried out to find out whether the bunching is effective. The signal received at 6.55 mm was plotted against the beam current, and it was found that the signal increases quadratically with beam current, thus indicating satisfactory bunching action. Measurements on the lines of shorter wave lengths were more difficult and it is hoped that further experiments will test the bunching in this region, too. It would be desirable to use the long undulator which has been constructed; i.e., the 3-foot model.

It should be mentioned that some lines in the submillimeter wave region were also observed when the electromagnet was switched off and completely demagnetised. Detailed interpretation of the observed spectrum has not been carried out so far. The work at Stanford had to be interrupted and it is hoped to continue it at Oxford. The distribution of the lines is puzzling and so is the appearance of lines in the absence of a magnetic field. The lines must of course be harmonics of the bunch repetition frequency. Explanation of these phenomena is being attempted along the following lines: It is possible that the surface properties of the guide through which the electrons pass play an important part. Suppose the wave guide walls are covered by a dielectric film. Eq. (90) can be simplified for a thin layer of dielectric of thickness $(b - a) = \Delta$. Then

$$I_o(x)/I_1(x) = \rho x$$
$$\rho = \Delta[\epsilon(v/c)^2 - 1]/a\epsilon[1 - (v/c)^2],$$

which has solutions

$$x = \xi_i;$$

e.g., for $x << 1$; i.e.,

$$[1 - (v/c)^2] << \Delta a\epsilon[\epsilon(v/c)^2 - 1]$$
$$\omega_r = v\xi_i/\rho = \epsilon[1 - (v/c^2]^{1/2}v/\Delta[\epsilon(v/c)^2 - 1].$$

A similar treatment can be applied to (77) for the corrugated guide showing the importance of surface irregularities of the guide. It is thus possible that some of the lines observed were in fact Cerenkov radiation. The influence of wall conductivity is also under investigation. Further experiments are necessary to clear up this matter.

An experiment on generation of Cerenkov radiation of K-band frequency (1.25 cm) was also carried out by M. Danos *et al.*[4] An electron beam accelerated by 10 kv (0.2 ma) was bunched by a K-band cavity and then made to graze a dielectric (polycrystalline T_iO_2) of dielectric constant $\epsilon = 105$. Approximately 10^{-7} watt of radiation was detected.

GENERAL CONSIDERATIONS

There has been a certain amount of discussion of the role of Cerenkov radiation in other microwave tubes such as klystrons and traveling-wave tubes. There is no doubt that the field radiated by electrons in slow wave structures is of the same nature as Cerenkov radiation discussed in previous sections. The bunching of the beam enhances the radiation. In traveling-wave tubes, however, the field emitted is an important feature of the dynamics of the electron itself. There is electron wave interaction. The bunching of the beam is affected by the radiated field. In resonant devices; i.e., in klystrons, the action of the fields on the electron

motion is equally important. The present paper does not consider electron-wave interaction but is confined to the primary process of radiation. The effect of bunching is an interaction effect of a special kind. If the undulator was made resonant the standing waves set up would in turn interact with the beam. But the frequencies radiated by high speed electrons due to the movement induced by these standing waves are much higher than those radiated due to the magnet structure. The frequencies at which power is abstracted from a highly bunched beam depend on the resonances of the cavity. The same is true of a resonant dielectric structure designed to extract Cerenkov radiation from bunched beams.

Nonreflecting Absorbers for Microwave Radiation

HANS SEVERIN†

Summary—The absorption of very short electromagnetic waves by absorbing systems, which avoid reflection of the incident wave is a problem of practical interest. Three different methods are applicable:

1. Complete absorption of the incident energy can be obtained for one wavelength by using resonance systems of relatively small thickness; e.g.,
a resistance card having a surface resistivity equal to the wave impedance of free space and placed a quarter of the wavelength in front of a metal sheet;
a dielectric layer of lossy material on a metal sheet, with the thickness of the layer equal to about a quarter of the wavelength in the material;
a two-dimensional periodic structure of concentric resonant circuits arranged within the metal sheet itself.

2. The reflecting object can be covered by a thick layer of absorbing material, so that in a wide wavelength range most energy of the incident wave will be absorbed before reaching the reflecting surface. To avoid reflection, the absorption material can be tapered or arranged in different layers in such a manner that the loss tangent steadily increases towards the base plate.

3. The bandwidth of resonance absorbers can be widened without an increase of its thickness by combination of two specially dimensioned resonant circuits.

IN CONNECTION with the rapid development of the technique of very short electromagnetic waves, a series of experimental and theoretical investigations has been carried out which treats certain devices absorbing electromagnetic radiation without producing much reflection.

If an electromagnetic wave is incident from a medium with the wave impedance Z_o upon a boundary surface characterized by the input impedance Z, the reflection coefficient r is given by the expression

$$r = \frac{Z - Z_o}{Z + Z_o}. \tag{1}$$

In free space the wave impedance is real $Z_o = X$ and for a plane electromagnetic wave

$$X = \sqrt{\frac{\mu_o}{\epsilon_o}} = 377\Omega \tag{2}$$

in all planes of constant phase. Therefore zero reflection can be obtained in a wide frequency band, when the input impedance Z of the absorber system is real and independent of frequency. Although the formulation of the problem is quite easy, it is difficult to realize the matching conditions in the microwave-range.

A first group of absorber types makes use of the resonance properties of systems having a relatively small layer thickness. The simplest and best-known device of that kind uses a resistance card whose surface resistivity is equal to the wave impedance of free space. If such a resistance foil is placed a quarter of the wavelength

from a metal plate, the input impedance in the foil plane is equal to the wave impedance of free space and perfect matching is achieved for the wavelength in question. The calculated wavelength sensitivity shows a voltage standing-wave ratio below 1.10, i.e. a reflection coefficient below 0.05, over a bandwidth of ±5 per cent.

Another resonance absorber of the same efficiency consists of a homogeneous dielectric layer on a metal sheet, where the layer thickness d, the dielectric constant ϵ, and the loss tangent $\tan \delta$ can be adjusted in such a way that reflection becomes zero for the desired wavelength.[1] The calculation of the reflection coefficient, based on Maxwell's equations, yields the expression

$$r = \frac{\tanh\left(2\pi i \frac{d}{\lambda} p\right) - p}{\tanh\left(2\pi i \frac{d}{\lambda} p\right) + p} \tag{3}$$

with

$$p = \sqrt{\epsilon(1 - i \tan \delta)}.$$

From the condition of zero reflection

$$p = \frac{\exp\left(2\pi i \frac{d}{\lambda} p\right) - \exp\left(2\pi i \frac{d}{\lambda} p\right)}{\exp\left(2\pi i \frac{d}{\lambda} p\right) + \exp\left(2\pi i \frac{d}{\lambda} p\right)} \tag{4}$$

follows as a first approximation for $2\pi d/\lambda < 1$, that the layer thickness d must be a quarter of the wavelength in the material

$$d = \frac{1}{n}\frac{\lambda}{4}, \qquad n = \sqrt{\frac{\epsilon}{2}\left(1 + \sqrt{1 + \tan^2 \delta}\right)}$$

and the surface resistivity has the value

$$\frac{1}{\sigma d} = \frac{1}{2} Z_o = 188{,}5 \ \Omega \text{ per square,}$$

(σ = conductivity of the layer material in Ω^{-1} cm^{-1}). In this case zero reflection arises from the interference of the two waves, which are reflected at the surface of the absorber and at the metal base respectively. The waves have a relative path difference of half a wavelength mainly determined by ϵ and d, and equal amplitudes determined by specially adjusted material losses, so that matching to the space is achieved.

With increasing wavelength the $\lambda/4$-condition of the two resonance systems described requires a considerable thickness of the absorber device. Whereas the resistance card must be located in a maximum of the electrical field strength, a magnetically operating absorber could be arranged immediately before a metal sheet. The

† III. Physikalisches Institut des Universitat Göttingen, Germany. At present at RCA Laboratories Division, Princeton, N. J.

[1] W. Dällenbach, and W. Kleinsteuber, "Reflection and Absorption of Decimeter-Waves by Plane Dielectric Layers," *Hochfreq. u Elektroak*, vol. 51, pp. 152–156; 1938.

prototype of an element reacting to the magnetic field is the magnetic dipole in the form of a loop antenna. Therefore a regular distribution of extremely small loop antennas on the metal sheet itself has been considered. The energy received from the field is led to load resistances. In such a manner we get an absorber device of small thickness, which is analogous to a resonance absorber of the Helmholtz type in the airborne sound technique.[2-4] In a first experimental investigation[5] within the range of decimeter waves small damped resonant circuits have been applied consisting of the inductance of the loop antenna, a series resistance and a tunable condenser. By arranging the loops before a metal plate the characteristic impedance of a waveguide could be reproduced and a matched termination realized. In the frequency range around 10,000 mc/s[6] a suitable resonator results from combining a capacitative coaxial line element with an inductive loop (Fig. 1). A few milli-
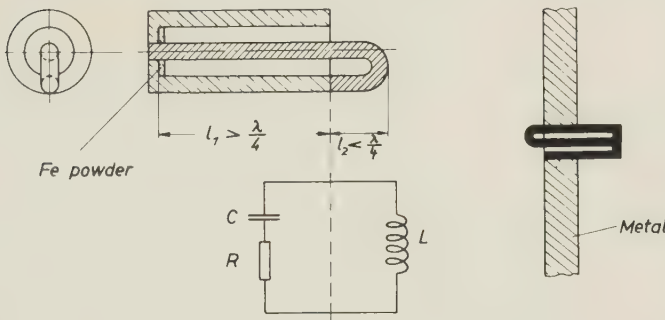
Fig. 1—Sectional view of a resonance element and its equivalent circuit.

grams of Fe-powder produce the required losses. Resonators tuned to the same resonance frequency are soldered in the metal sheet. The loop antennas are in parallel with each other and form a two-dimensional periodic structure. To get minimum reflection two parameters can be varied, namely the number of resonance elements per unit area and the resistance of the resonators. Fig. 2 shows some results of measurements on the influence of these two parameters. Optimum matching is achieved for a certain number of resonance elements per unit area. If the number is too small, the input impedance of the absorber is lower than the characteristic impedance of free space. On the other hand if the surface density of elements exceeds its optimum value, the input impedance becomes too high and the reflection coefficient increases again. Because the input impedance

[2] Resonators in the form of cavities with necks are arranged in a rigid wall and the kinetic energy effective in the necks is dissipated by a suitable flow resistance.[3,4]

[3] W. Willms, "On sound absorption with aid of damped resonators," Akus. Zeit., vol. 4, pp. 29–32; 1939.

[4] H. Oberst, "Resonant absorbers for air-borne sound," Rep. III Physikalisches Institut der Universität, Göttingen, Germany; 1952. Not published up to now.

[5] E. Meyer, and W. Lippert. Report. Heinrich-Hertz-Institut für Schwingungsforschung. Berlin, Germany; 1944. Not published.

[6] E. Meyer, H. Severin and G. Umlauft. "Resonance absorbers or electromagnetic waves," Z. Phys., vol. 138, pp. 465–477; 1954.

also increases with the resistance of the resonators, matching is obtained with the smaller optimum surface density ($N = 1.12$ cm^{-2}) when the elements with the higher resistance (1.6 mg Fe-powder) are used.

The trend of the reflection coefficient as a function of the wavelength and the number of elements per unit area will be the same if the $\vec{E}$-vector of the incident plane wave is parallel or perpendicular to the plane of incidence. The dependence of the reflection coefficient on the angle of incidence, however, must be expected to be different for the two polarizations. In the first case, the $\vec{H}$-vector is perpendicular to the loop areas, which remain effective independent of the angle of incidence.
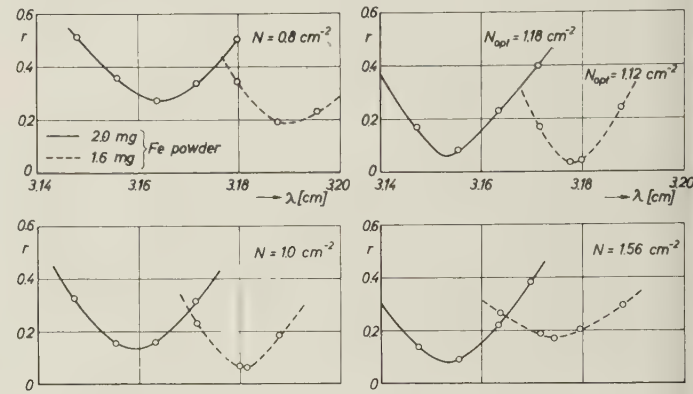

Fig. 2—Reflection coefficient of resonance absorber sheets as a function of the wavelength, measured for polarization perpendicular to the plane of incidence, angle of incidence 15 degrees. Parameters: Number per cm^2 and resistance of the absorbing elements.

Therefore, the surface impedance too is independent of the angle of incidence. For the other polarization only a component of the magnetic field has the direction of the loop normals. With increasing angle of incidence the effective loop area and with it the surface impedance are diminished. Together with the characteristic impedance of free space

$$Z_o = X \cos \varphi, \qquad Z_o = \frac{X}{\cos \varphi} \qquad \text{(5a, b)}$$

we get from (1) for $\vec{E^i}$ parallel to plane of incidence

$$r = \frac{Z - X \cos \varphi}{Z + X \cos \varphi} \qquad \text{(6)}$$

$\vec{E^i}$ perpendicular to plane of incidence

$$r = \frac{Z(\varphi) \cos \varphi - X}{Z(\varphi) \cos \varphi + X}. \qquad \text{(7)}$$

Fig. 3 shows the angle dependence of the reflection coefficient for the two optimum absorbing sheets of Fig. 2. For a vertically incident wave the absorber with 1.12 elements per centimeter square has an input impedance higher than $X = 377\Omega$. Therefore, according to (7), in the case of vertical polarization the reflection coefficient decreases with increasing angle of incidence, has a minimum for $Z(\varphi) \cos \varphi \approx X$, and increases afterwards.

n the other case of parallel polarization according to
), the reflection coefficient increases monotonic
ith the angle of incidence. The input impedance of
he other absorber having 1.18 elements per centimeter
quare is lower than $X = 377\ \Omega$ at the angle of incidence
$= 0$ degrees. Therefore in this case, according to (6)
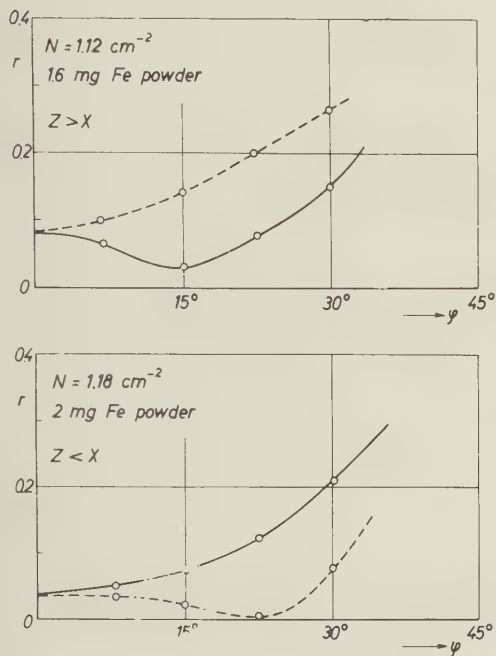nd (7), the behavior is opposite for the two polariza-
ions.

Fig. 3—Reflection coefficient of resonance absorber sheets as a
function of angle of incidence, measured for polarization per-
pendicular (—) or parallel (---) to plane of incidence. ($Z =$ input
impedance of the absorber sheet at $\varphi = 0$ degrees, $X = 377\Omega$.)

The absorber described is effective only when the
magnetic vector of the incident wave has a component
in the direction of the magnetic dipoles represented by
the loop antennas. Practice, however, demands the
absorption be independent of the arbitrary angle between
the magnetic vector and the axis of the dipoles. That
can be achieved by adding a second grating of resonance
elements, the loop antennas of which are turned 90
degrees with respect to those of the first one. Because
of the considerable manufacturing difficulties in the
range of centimeter waves, the field of practical appli-
cation of this type of resonance absorber will be in the
range of decimeter- and meter-waves.

The second group of nonreflecting microwave ab-
sorbers is characterized by a relatively large effective
bandwidth, and as a consequence of this property by a
large layer thickness. In contrast to the resonance sys-
tems described above, most of the energy of the incident
wave will be absorbed before reaching the reflecting
surface and multiple reflections in the layer itself are
avoided. Assuming this, a formal solution of the prob-
lem is quite easy. We need a material, the characteris-
tic impedance of which is equal to that of free space:

$$Z = \sqrt{\frac{\mu_o}{\epsilon_o}}\sqrt{\frac{\mu(1 - \tan \delta_\mu)}{\epsilon(1 - \tan \delta_\epsilon)}}$$

$$= \sqrt{\frac{\mu_o}{\epsilon_o}}\sqrt{\frac{\mu}{\epsilon}\frac{\cos \delta_\epsilon}{\cos \delta_\mu}} \cdot \quad e^{\frac{-i-\delta_\mu-\delta_\epsilon}{2}} = \sqrt{\frac{\mu_o}{\epsilon_o}} = X. \quad (8)$$

An efficient absorber should have large and equal values
of the dielectric and magnetic loss angles δ_ϵ and δ_μ as
well as of dielectric constant ϵ and permeability μ, and,
moreover, these material constants should be inde-
pendent of frequency. The realization of an absorber
utilizing these principles is not possible up to now.

In the actual construction of a wide band absorber for
microwaves, reflection is avoided by tapering a homoge-
neous absorption material, or by arranging different lay-
ers parallel to the base plate in such a manner that the
loss tangent increases towards the base plate. The special
problem of two conducting layers and two equal air gaps
in front of a metal sheet is exactly treated by solving
Maxwell's equations.[7] By this calculation it is found that
two zeroes of the reflection coefficient now exist. Numeri-
cal evaluations of the results show the possibility of
reducing the reflection coefficient below 0.1 within the
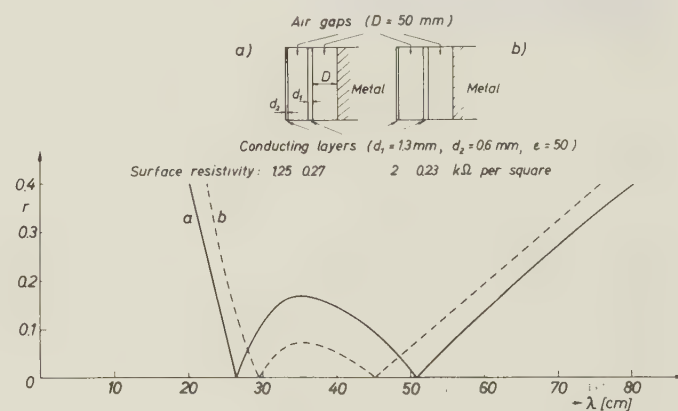range of one octave (see Fig. 4). In another construction

Fig. 4—Calculated reflection coefficient of a two-layer absorber as a
function of the wavelength.

the absorber consists of alternate layers of a lossless
dielectric and thin sheets of poorly conducting material,
the surface resistivity of which decreases towards the
base metal plate by a constant factor from one sheet
to the next. The behavior of this so called "electrical
swamp" may be approximately calculated by means of
the transmission line theory.[8] Fig. 5 shows the reflection
coefficient calculated in this way as a function of wave-
length to be below 0.1 in a range of nearly 3 octaves.
The actual absorbers, however, do not give as good
results as calculated because of variations in the various

[7] R. Becker, "The reflection of electromagnetic waves at a
medium consisting of layers," Rep. Institut für Theoretische
Physik der Universität, Göttingen, Germany; 1943. Not published.
[8] C. G. Montgomery, R. H. Dicke and E. M. Purcell, "Principles
of Microwave Circuits," M.I.T. Rad. Lab. Ser., vol. 8, McGraw-
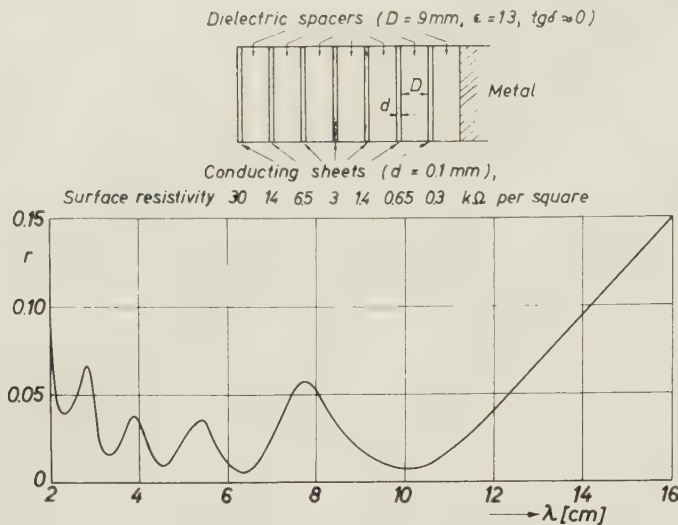Hill Book Co., New York, N. Y., p. 396–399; 1948.

Fig. 5—Calculated reflection coefficient of a multilayer absorber as a function of the wavelength.

Fig. 6—Attenuation of an electromagnetic wave (λ = 3.2 cm) in a medium of parallel conducting plates as a function of the surface resistivity, for different spacings *d*.

Fig. 7—Specular reflection-wavelength characteristic for an absorber made of indented resistance foils.

parameters in manufacture. The reflection coefficient varies from 0.05 to 0.1 in the wavelength range of 7 to 15 cm.

The manufacturing difficulties arising from the exact observance of the calculated values of the surface resistivity can be avoided by using a uniform absorbing material in such a geometric shape that the absorption coefficient increases towards the base plate. This method of the gradual transition using wedge-shaped, pyramidal or conical structures is well-known from matched transmission-line terminations[9] and from the airborne[10] and the waterborne[11] sound techniques. A special construction of this electrical absorber is built up with resistance cards arranged perpendicular to the base plate and parallel to the direction of the electric field of the incident wave. To determine the optimum dimensioning of this device with respect to the surface resistivity and the layer distance, the attenuation of an electromagnetic wave in a medium of parallel conducting plates has been investigated (Fig. 6).[12] For spacings smaller than half a wavelength and small surface resistivities, the behavior is quite similar to that of a waveguide below the cutoff frequency. High damping values are achieved because waves cannot be propagated and the field decays exponentially along the guide. Consequently the incident wave does not penetrate into the parallel plate medium and high reflection must be expected. For our special purpose, therefore, systems with spacings greater than half a wavelength will be applied. Then, between the two limiting cases of lossless wave propagation in perfectly conducting guides and in
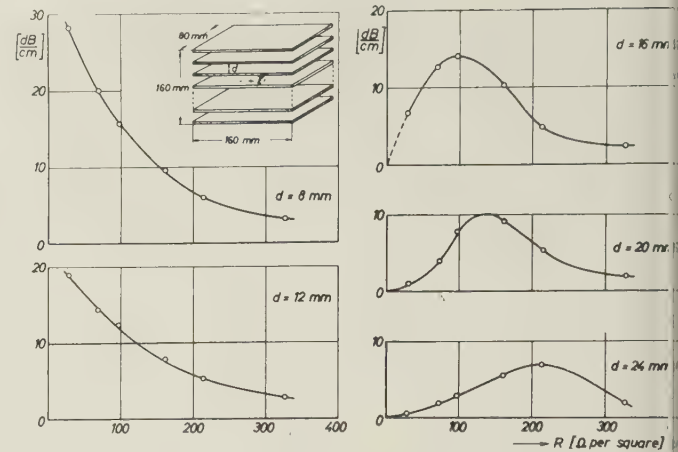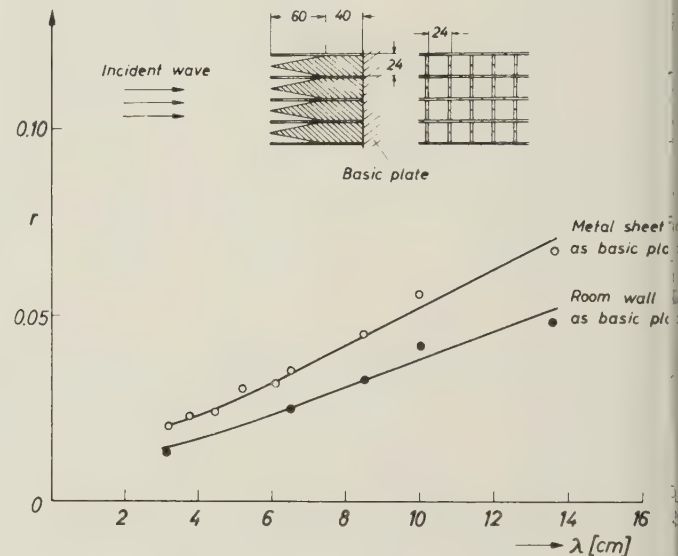
free space, characterized by zero- and infinite surface resistivity, respectively there is a maximum of attenuation per unit length, which is shifted to higher values of surface resistivity with increasing spacing. For practical applications, spacings and corresponding surface resistivities for maximum attenuation are chosen in such a way that the absorber thickness necessary for getting the required total absorption is not too large. According to Fig. 6 the suitable distance of the conducting sheet is a little larger than half the wavelength. In the special case of λ = 3.2 cm the experimental results yield *d* = 2.0 cm, *R* = 150 Ω per square and α = 10 *dB/cm*.

To match the absorbing parallel plate medium to the characteristic impedance of free space, the conducting sheets are indented on the side of the incident wave. To avoid a dependence on polarization, two such systems are arranged on the base plate perpendicular to it and to each other. An absorber of this kind of the size 1 *m²* has been manufactured and showed a reflection coefficient *r* < 0.1 in the range of at least three octaves (Fig. 7). The angle dependence of the

[9] C. G. Montgomery, "Technique of Microwave Measurements," M.I.T. Rad. Lab. Ser., vol. 11, McGraw-Hill Book Co., Inc., New York, N. Y., p. 720–743; 1947.

[10] E. Meyer, G. Buchmann and A. Schoch, "A New Sound Absorbing Arrangement of High Efficiency and the Construction of an Anechoic Chamber," *Akus. Zeit.*, vol. 5, pp. 352–364; 1940.

[11] E. Meyer, and K. Tamm, "A wide-band absorber for waterborne sound," *Acustica*, vol. 2,; AB 91–104; 1952.

[12] H. Sauer, Diplomarbeit, III. Physikalisches Institut der Universität, Göttingen, Germany; 1955.
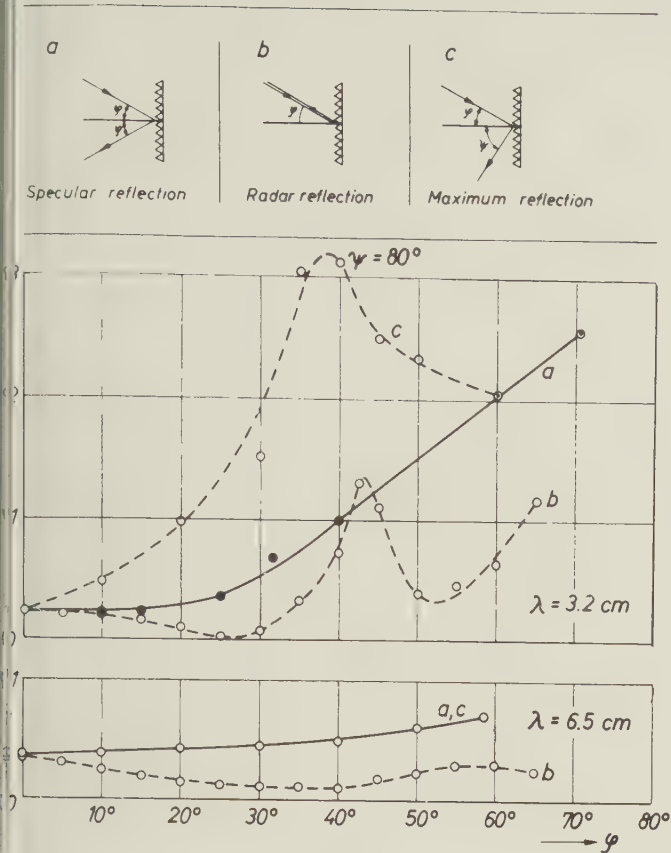
a   *b*   *c*

Specular reflection   Radar reflection   Maximum reflection

Fig. 8—Reflection of an absorber made of indented resistance foils as a function of the angle of incidence.

Fig. 9—Weather-proof wide-band absorber.

Fig. 10—Specular reflection-wavelength characteristic for flat material (---) and pyramid material (—).

flection coefficient of the same absorber is measured under various conditions of practical interest and the results are reproduced in Fig. 8. An interference effect resulting in an intense reflection can be observed. This ffect arises from the periodic structure of the absorber, the spacings are larger than half the wavelength, and an be avoided by using teeth of different lengths. This rinciple can be seen in Fig. 9 showing a weatherproof bsorber,[13] where the conducting sheets are embedded a foamed dielectric ($\epsilon \approx 1.05$).

Of course, instead of indented foils, other forms of apering can be used such as wedges, pyramids or cones. ither the surface of these structures can be covered ith a thin conducting film, or they can be formed from homogeneous lossy material. Some applications of these ide-band absorbers are radar camouflage of disturbing uperstructures on air fields or ships, the decoupling of ransmitting and receiving antennas of a microwave elay station, and, for measuring purposes, the coating of oom walls on which a directed beam is incident, or the onstruction of a free-space room for electromagnetic aves. Such rooms can be used for indoor investigations f wave propagation phenomena, as the radiation patterns f directional antennas, the diffraction at openings and e back-scattering from obstacles. The Naval Research
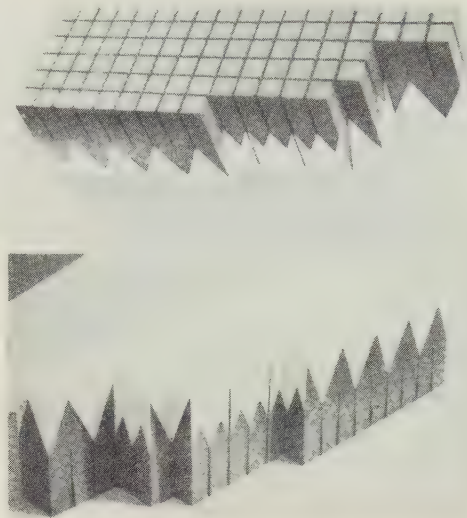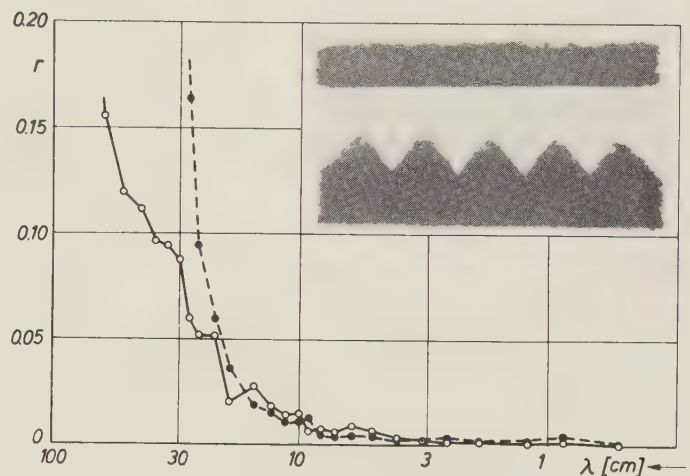
Laboratory at Washington has constructed such an anechoic chamber for microwaves in the region of $\lambda = 1.2$ to $\lambda = 12$ cm.[14] The nonreflecting material was made by impregnating commercially available mats of curled animal hair with a mixture of conducting carbon black in neoprene. Reflection from the interface has been minimized by either geometrical shaping of the front surface or by increasing the loss with depth. Fig. 10 shows specular reflection as a function of the wavelength at angles near normal incidence for both types of materials. The III. Physical Institute of the University at Goettingen has constructed a large free-space room, which is anechoic for both acoustic and electromagnetic waves.[15] The walls are lined with sound-absorbing glass-wool wedges, which are made electrically absorbent by

[13] Manufacturer: W. Genest, Stuttgart-Degerloch, Germany.

[14] A. J. Simmons, and W. H. Emerson, "Anechoic Chamber for Microwaves," *Tele-Tech. and Electronic Industries*, pp. 47–49; July, 1953.
[15] E. Meyer, G. Kurtze, H. Severin and K. Tamm, "A New Large Free Space Room for Acoustic and Short Electromagnetic Waves," *Acustica*, vol. 3, pp. 409–420; 1953.
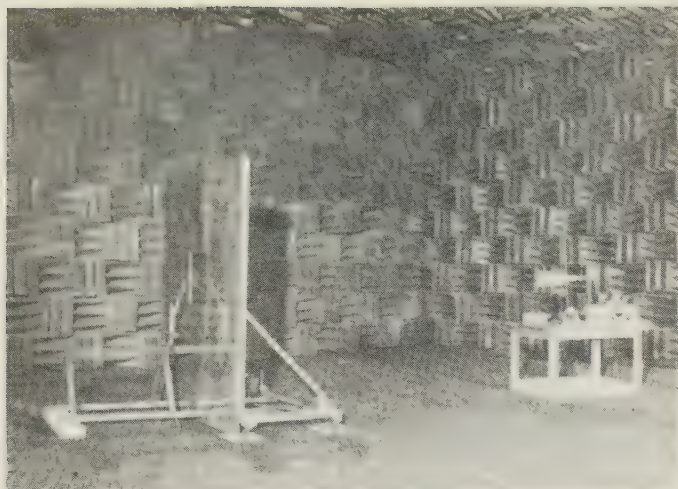
Fig. 11—Anechoic chamber for acoustic and electromagnetic waves. III. Physical Institute of the University at Goettingen.

The disadvantage of these wide-band absorbers is th great thickness of the whole system, which does no allow their technical application to movable object The practical use of the resonance absorbers, on th other hand, is reduced by their small bandwidth. It i however possible to increase the effective bandwidth b combining different circuits. In low-frequency techniqu it is possible to construct a two-terminal network wit an input impedance, which is real and independent c frequency by combining two specially dimensioned resc nant circuits (see Fig. 13). A network of this kind whic has already been used very successfully with purely me chanical elements in the water borne sound technique can also be realized in the microwave-range.[17] By regula distribution of such elements on a surface, an absorbe system with defined properties, determined by the singl absorber elements and their distribution, can be obtainec
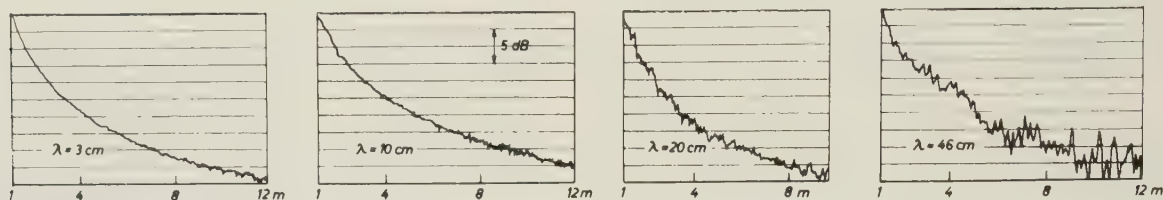
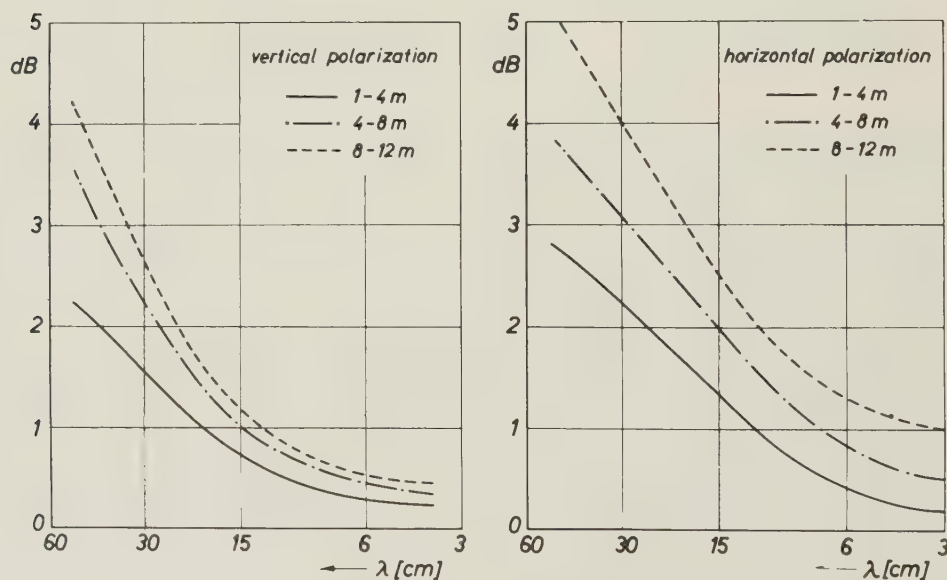Electric field strength as a function of the distance from the dipol (vertical polarization).

Fig. 12—Maximum deviations of the 1/r-law of propagation of an electric dipole for different distances from the dipole. Polarization is referred to the steel net floor.

sucking in graphite powder (Fig. 11). The reflection coefficient of this arrangement is below 0.1 for electromagnetic waves below 30 cm wavelength. The quality of the finished room was tested by measuring the deviations of the 1/r law of propagation of dipole-radiation and the results are reproduced in Fig. 12.

A short-circuited line, a quarter of a wavelength long represents a parallel resonant circuit in the vicinity of its resonant frequency, so that, in order to make a meta

[16] E. Meyer, and H. Oberst, "Resonant Absorbers for Water-borne Sound," Acustica, vol. 2, AB 149–170; 1952.
[17] H. J. Schmitt, Dissertation at III. Physikalisches Institut der Universität, Göttingen, Germany; 1955.
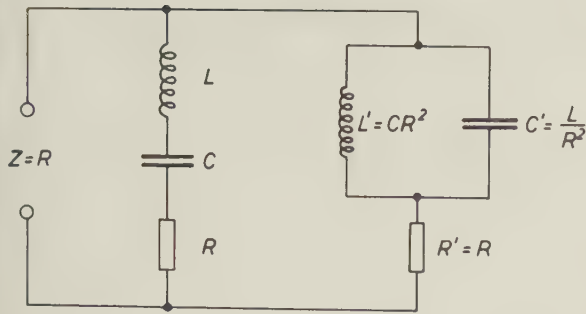
Fig. 13—Two-terminal network of real input impedance, independent of frequency.

plate nonreflecting, suitable elements of series circuit character have to be mounted at a distance of λ/4 in front of the metal plate. Among such elements are electric dipoles having the same resonant frequency as the parallel resonant circuit. The necessary damping of the dipoles can be obtained by manufacturing them of resistive material. For some applications of these absorbers it is of special interest to reduce the total depth of the device. For this purpose the λ/4 spacing can be reduced to $\lambda/4\sqrt{\epsilon}$ by using a material with the dielectric constant ε.
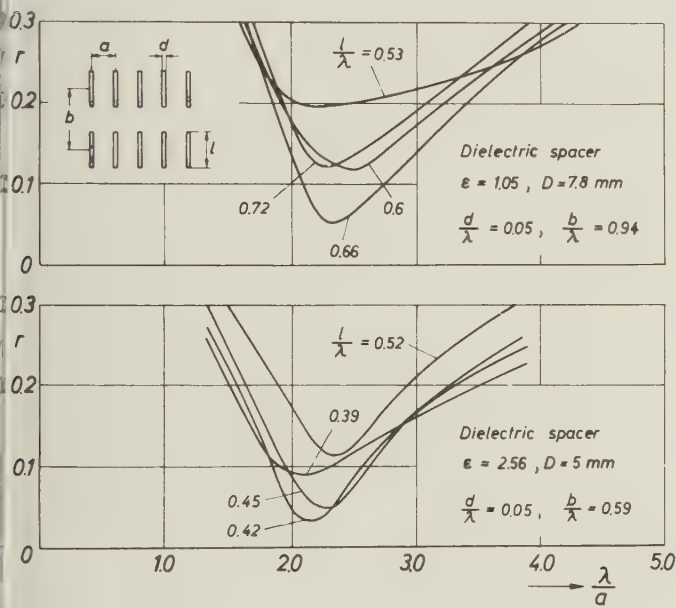

Fig. 14—Reflection coefficient of dipole absorbers as a function of the grating constant for different lengths of the dipoles. λ = 3.2 cm, surface resistivity of the dipoles: 40 Ω per square.

To determine optimum dimensions the reflection coefficient of absorber plates was measured for different lengths of the dipoles and different grating constants. The wavelength was 3.2 cm and the dipoles were parallel to the electric vector of the incident wave. Fig. 14 shows that for each length of the dipoles there is a grating constant for which the reflection coefficient is minimum. At the resonance-length this minimum

reaches very small values, corresponding to an input impedance of the absorber of nearly $Z_o = 377\ \Omega$. At an absorber with dielectric spacer the absolute minimum of the reflection coefficient is reached for substantially smaller (by the factor $\sqrt{\epsilon}$) resonance lengths of the dipoles. In Fig. 15 the measured dependence of the re-

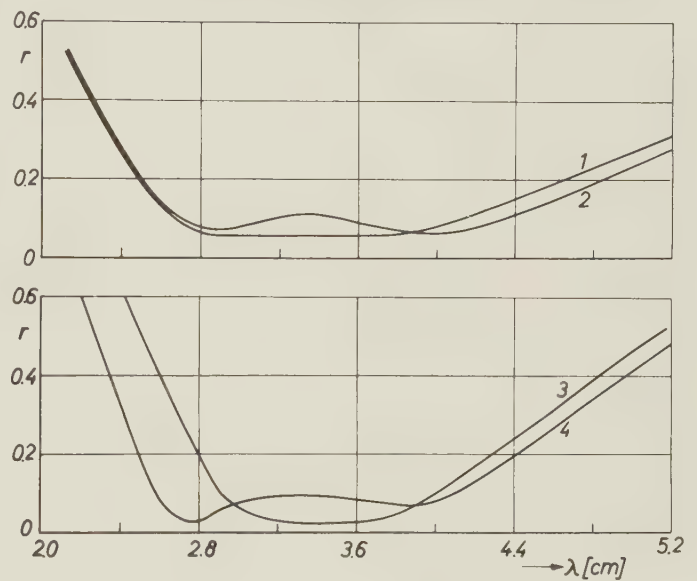| absorber | spacer | | $\frac{d}{\lambda}$ | $\frac{l}{\lambda}$ | $\frac{a}{\lambda}$ | $\frac{b}{\lambda}$ |
	ε	D				
1	}1.05	7.8 mm	0.05	0.66	0.39	}0.94
2			0.03	0.72	0.29	
3	}2.56	5.0 mm	}0.05	}0.42	0.47	}0.59
4					0.39	

Fig. 15—Reflection coefficient of various dipole absorbers as a function of the wavelength.

flection coefficient of various absorber plates on wavelength is given. With an upper limit of 0.1 for the reflection coefficient the effective bandwidth extends in the best case from λ = 2.6 cm to λ = 4.1 cm with an absorber system having a total depth of 5 mm. With a wavelength λ = 1.6 cm we have metallic reflection since in this case the dipoles are placed in a node plane of the electric field in front of the metal plate. Finally, Fig. 16 shows the dependence of the reflection coefficient of an absorber plate on the angle of incidence for the cases in which the electric vector of the incident wave is parallel or perpendicular to the plane of incidence. In the latter case the angle dependence can be calculated, and the theoretical curve for an air spacing has been drawn for comparison. With a dielectric spacer the reflection coefficient for oblique incidence is always smaller than with an air spacing. In the dielectric layer the incident wave is refracted towards the normal of the layer. Hence for oblique incidence the maximum of the electric field shifts less as the dielectric constant becomes
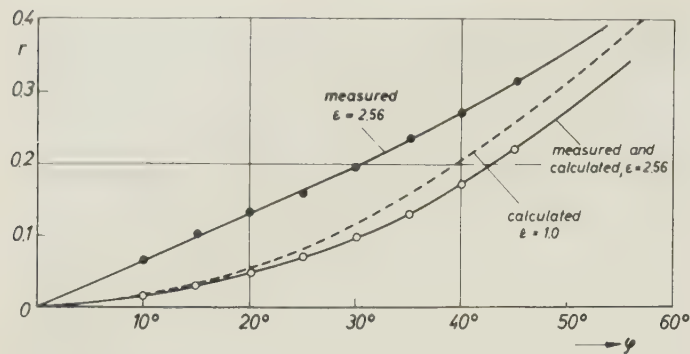
Fig. 16—Reflection coefficient of a dipole absorber with plexiglass dielectric, matched for normal incidence as a function of the angle of incidence. Polarization parallel (•—•) or perpendicular (○—○) to the plane of incidence.

higher. In the form described the absorber is effective only when the electric vector of the incident wave has a component in the direction of the dipole axis. Practice, however, demands very small reflection coefficients independent of the polarization. This can be achieved by adding a second grating, the elements of which are turned by 90 degrees with respect those of the first one, i.e. a grating of crosses instead of dipoles. Absorber-plates of this kind have been constructed, which, as expected, are practically independent of the polarization and are equally effective as the dipole systems. According to the results reproduced in Figs. 15 and 16 the main field of application of these absorbers will be radar camouflage when the transmitter frequencies are known.

Theory of the Corner-Driven Square Loop Antenna

RONOLD KING†

Summary—The general problem of determining the distribution of current and the driving point impedances of a square loop or frame antenna is formulated when arbitrary driving voltages are applied at each corner or when up to three of these voltages are replaced by impedances. The loop is unrestricted in size and account is taken of the finite cross-section of the conductors. Four simultaneous integral equations are obtained and then replaced by four independent integral equations using the method of symmetrical components. These equations are solved individually by iteration and first-order formulas are obtained for the distributions of current and the driving-point admittances. By superposition the general solution for the abritrarily driven and loaded loop is obtained. Interesting special cases include a corner-reflector antenna and the square rhombic (terminated) antenna. An application of the principle of complementarity permits the generalization of the solution to the square slot antenna in a conducting plane when driven from a double-slot transmission line at one corner.

General Formulation

THE SQUARE loop antenna driven at one or more corners and loaded at others has many applications ranging from beacons and television transmitters to direction finders. Whereas the circuit and radiation characteristics of small loops are well-known,[1, 2] the square loop that is not small compared with the wavelength has not been studied systematically. A rigorous analysis of the current and impedance of the circular loop when driven at one point was made by Hallén.[3] Since the solution was not convenient for numerical evaluation it has been re-examined by Storer[4] and modified to facilitate computation. Extensive measurements of impedances and distributions of current for both square and circular loops have been made by Kennedy,[5] using an image-plane and by Kaliszewski, using complementary slots.[6]

The circuit to be analyzed is the square loop shown in Fig. 1(a). The sides of the square (numbered 1 to 4) are constructed of highly-conducting wire and the loop is driven at each corner from a two-wire line. Alternative connections are illustrated in Figs. 1(b), 2, and 3. The complementary slot circuit shown in Fig. 4 is useful for exciting surface waves on conducting planes. A shielded loop is shown in Fig. 5. A square loop driven at one corner and loaded at the diagonal corner is the 90 degrees rhombic antenna shown in Fig. 6.

Throughout the following analysis it is assumed that *only the currents in the loop itself contribute to radiation.*

† Craft Lab., Harvard Univ., Cambridge, Mass.
[1] R. W. P. King, "Electromagnetic Engineering," vol. I, McGraw-Hill Book Co., Inc., New York, N. Y., pp. 427 ff; 1945.
[2] R. King, "A loop transmitter," *Phil. Mag.*, ser. 7, vol. 20, pp. 514–528; 1935.
[3] E. Hallén, "Theoretical investigations into transmitting and receiving antennae," *Nova Acta (Uppsala)*, ser. 4, vol. 11, pp. 1–44; 1938.
[4] J. E. Storer, Cruft Lab. Tech. Rep. No. 212, Harvard Univ., Cambridge, Mass.
[5] P. A. Kennedy, "Loop Antenna Measurements," Cruft Lab. Tech. Rep. No. 213, Harvard Univ., Cambridge, Mass.
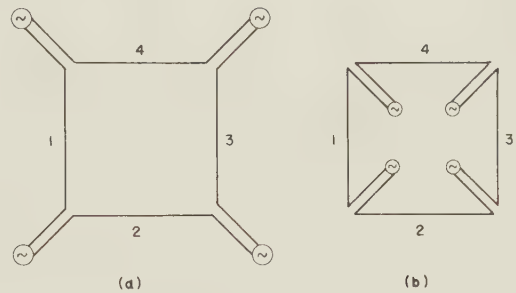[6] T. Kaliszewski, unpublished research.

Fig. 1.—Corner-driven loop; transmission lines in the plane of the loop.
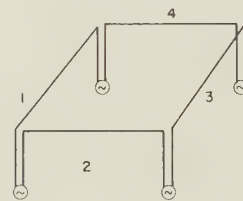
Fig. 2.—Corner-driven loop; transmission lines perpendicular to the plane of the loop.
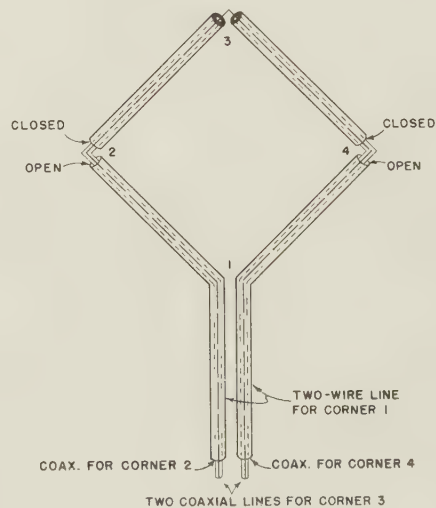
Fig. 3.—Corner-driven square loop with internal feed lines.

Therefore, none of the lines may be unbalanced. *Circuits that do not satisfy this condition are excluded.*

It is shown in the literature[7] with particular reference to excitation with a two-wire-line that transmission-line end effects and coupling between the transmission lines and the antenna are different in the two circuits illustrated in Figs. 1(a) and 2, and it is readily argued that Fig. 1(b) presents yet a third case.

In all three the measurable apparent impedances terminating the several transmission lines differ not

[7] R. King, "Theory of V-antenna," *Jour. Appl. Phys.*, vol. 22, pp. 1111–1121; 1951.

Fig. 4.—Complementary-slot square loop.

Fig. 5.—Shielded square loop.

Fig. 6.—Ninety-degree rhombic antenna.

only from one another but also from the corresponding ideal values obtained when each corner is driven by a lumped impedanceless generator. However, this ideal case may be modified to approximate closely an actual driving line by introducing an equivalent lumped network at each corner in conjunction with a concentrated emf as shown in F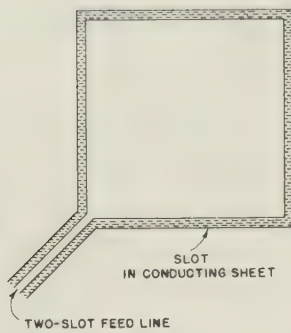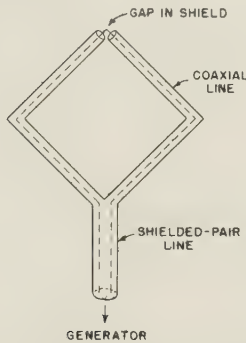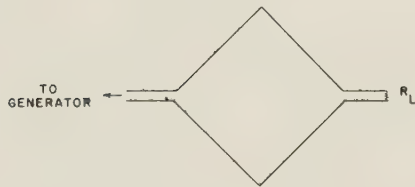ig. 7. The constants of the lumped network depend upon the particular arrangement and spacing of the driving lines. Their determination is not the primary problem of this investigation. Methods are described in the literature.[7-10] For present purposes it is sufficient to treat the circuit in Fig. 4 in which the precise nature of the generators and networks at the corners is not specified, and it is assumed that the voltages V_{12}, V_{23}, V_{34}, and V_{41} are maintained across the adjacent ends of the cylindrical conductors forming

[8] R. King and K. Tomiyasu, "Terminal impedance and generalized two-wire line theory," Proc IRE, vol. 37, pp. 1134–1139; October, 1949.

[9] R. W. P. King, "Transmission-Line Theory," McGraw-Hill Book Co., Inc., New York, N. Y., Chap. V; 1955.

[10] R. King, "Antennas and open-wire lines, part I," Jour. Appl. Phys., vol. 20, pp. 832–850; 1949.

the sides of the square (dimension $2h$). These are numbered from 1 to 4 as shown in Fig. 7 and each has a length $2(h - \delta)$ and a radius a. It is assumed that the following inequalities are satisfied:

$$a \ll h; \qquad \beta_o a \ll 1; \qquad (1)$$

$$\delta \ll h; \qquad \beta_o \delta \ll 1; \qquad (2)$$

where $\beta_o = 2\pi/\lambda_o$. The origin of a right-handed rectangular coordinate system is at the center of the square.
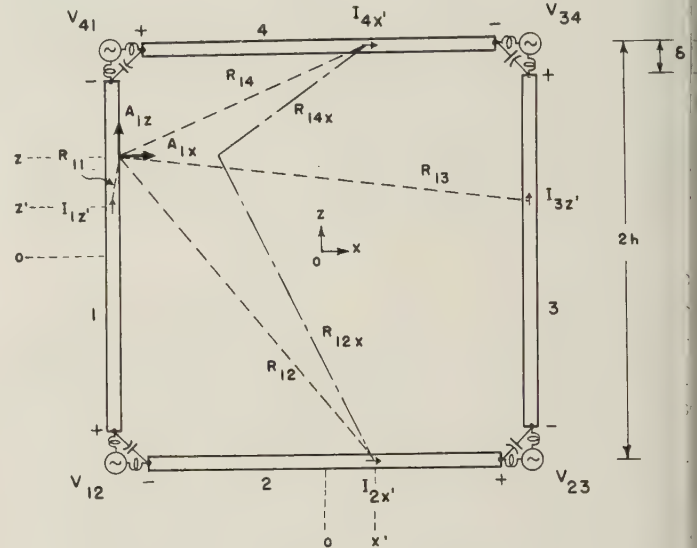

Fig. 7.—Square loop with lumped emf and junction network at each corner.

The required vanishing of the tangential component of the electric field on the surface of each conductor leads to four equations of which that for conductor 1 is typical. This is

$$-E_{1z} = \frac{\partial \phi_1}{\partial z} + j\omega A_{1z} = \frac{\partial \phi_{1x}}{\partial z} + \frac{\partial \phi_{1z}}{\partial z} + j\omega A_{1z} = 0, \quad (3)$$

where

$$\phi_{1x} = \left[\frac{j\omega}{\beta_o^2} \frac{\partial A_{1x}}{\partial x} \right]_{x=-h}; \quad \phi_{1z} = \left[\frac{j\omega}{\beta_o^2} \frac{\partial A_{1z}}{\partial z} \right]_{x=-h}. \quad (4)$$

[Note that the subscripts x and z denote Cartesian components of the vector potential A and parts of the scalar potential ϕ associated with these components according to (4)]. The scalar and vector potentials are defined on the surface of conductor 1. The components of the vector potential are given by the following integrals. Note that $\nu_o = 1/\mu_o$.

$$A_{1z}(z) = \frac{1}{4\pi\nu_o} \int_{-h+\delta}^{h-\delta} \left[I_{1z}(z') \frac{e^{-j\beta_o R_{11}}}{R_{11}} \right.$$
$$\left. + I_{3z}(z') \frac{e^{-j\beta_o R_{13}}}{R_{13}} \right] dz'; \quad (5)$$

$$A_{1x}(z) = \frac{1}{4\pi\nu_o} \int_{-h+\delta}^{h-\delta} \left[I_{2x}(x') \frac{e^{-j\beta_o R_{12}}}{R_{12}} \right.$$
$$\left. + I_{4x}(x') \frac{e^{-j\beta_o R_{14}}}{R_{14}} \right] dx'; \quad (6)$$

where

$$R_{11} = \sqrt{(z - z')^2 + a^2}; \quad R_{13} = \sqrt{(z - z')^2 + (2h)^2}; \quad (7)$$

$$R_{12} = \sqrt{(h + z)^2 + (h + x')^2 + a^2};$$

$$R_{14} = \sqrt{(h - z)^2 + (h + x')^2 + a^2}. \quad (8)$$

The scalar potential on conductor 1, $\phi_1 = \phi_{1x} + \phi_{1z}$, may be expressed in terms of currents using (5) and (6) with (4), or it may be expressed in terms of the charge per unit length as follows:

$$\phi_{1z}(z) = \frac{1}{4\pi\epsilon_o} \int_{-h+\delta}^{h-\delta} \left[q_1(z') \frac{e^{-j\beta_0 R_{11}}}{R_{11}} \right.$$

$$\left. + q_3(z') \frac{e^{-j\beta_0 R_{13}}}{R_{13}} \right] dz'; \quad (9)$$

$$\phi_{1x}(z) = \frac{1}{4\pi\epsilon_o} \int_{-h+\delta}^{h-\delta} \left[q_2(x') \frac{e^{-j\beta_0 R_{12}}}{R_{12}} \right.$$

$$\left. + q_4(x') \frac{e^{-j\beta_0 R_{14}}}{R_{14}} \right] dx'. \quad (10)$$

Note that $\phi_{1z}(z)$ is the potential at a point z on conductor 1 due to the charge distributions $q_1(z')$ and $q_3(z')$ on the two conductors that are parallel to the z-axis. Similarly, $\phi_{1x}(z)$ is the potential at z on conductor 1 due to the charge distributions $q_2(x')$ and $q_4(x')$ on the two conductors that are parallel to the x-axis.

With (4) substituted in (3) the following differential equation is obtained for the z-component of the vector potential on the surface of conductor 1:

$$\frac{\partial^2 A_{1z}}{\partial z^2} + \beta_o^2 A_{1z} = j \frac{\beta_o^2}{\omega} \frac{\partial \phi_{1x}}{\partial z}. \quad (11)$$

By differentiating (11) with respect to z and making use of (4) the following equation is obtained for the total scalar potential $\phi_1(z) = \phi_{1x}(z) + \phi_{1z}$ at z, on conductor 1:

$$\frac{\partial^2 \phi_1(z)}{\partial z^2} + \beta_o^2 \phi_1(z) = \beta_o^2 \phi_{1x}(z). \quad (12)$$

The solution of (11) is the sum of a complementary function and a particular integral. Thus,

$$A_{1z}(z) = \frac{-j}{v_o} [C_1 \cos \beta_o z + C_2 \sin \beta_o z - \theta_A(z)], \quad (13)$$

where $v_o = 1/\sqrt{\mu_o \epsilon_o}$ and where

$$\theta_A(z) = \int_{z_0}^{z} \frac{\partial \phi_{1x}(w)}{\partial w} \sin \beta_o(z - w) \, dw. \quad (14)$$

The lower limit z_o in (14) is arbitrary. An alternative form of (14) is obtained with integration by parts. Thus,

$$\theta_A(z) = \beta_o \int_{z_0}^{z} \phi_{1x}(w) \cos \beta_o(z - w) \, dw$$

$$- \phi_{1x}(z_o) \sin \beta_o(z - z_o). \quad (15)$$

With the notation,

$$\theta_c(z) = \int_{z_0}^{z} \phi_{1x}(w) \cos \beta_o w \beta_o \, dw;$$

$$\theta_s(z) = \int_{z_0}^{z} \phi_{1x}(w) \sin \beta_o w \beta_o \, dw, \quad (16)$$

$\theta_A(z)$ as given in (15) becomes:

$$\theta_A(z) = [\theta_c(z) + \phi_{1x}(z_o) \sin \beta_o z_o] \cos \beta_o z$$

$$+ [\theta_s(z) - \phi_{1x}(z_o) \cos \beta_o z_o] \sin \beta_o z. \quad (17)$$

The solution of (12) is conveniently obtained from (13) using (24). Thus,

$$\dot{\phi}_1(z) = \frac{jw}{\beta_o^2} \frac{\partial A_{1z}(z)}{\partial z} + \phi_{1x}(z)$$

$$= -C_1 \sin \beta_o z + C_2 \cos \beta_o z - \frac{1}{\beta_o} \frac{\partial \theta_A(z)}{\partial z} + \phi_{1x}(z). \quad (18)$$

With (17), it follows that

$$\theta_V(z) = -\frac{1}{\beta_o} \frac{\partial \theta_A(z)}{\partial z} + \phi_{1x}(z)$$

$$= [\theta_c(z) + \phi_{1x}(z_o) \sin \beta_o z_o] \sin \beta_o z - [\theta_s(z) - \phi_{1x}(z_o)$$

$$\cos \beta_o z_o] \cos \beta_o z = \int_{z_0}^{z} \phi_{1x}(w) \sin \beta_o(z - w) \beta_o \, dw$$

$$+ \phi_{1x}(z_o) \cos \beta_o(z - z_o). \quad (19)$$

Hence,

$$\phi_1(z) = -C_1 \sin \beta_o z + C_2 \cos \beta_o z + \theta_V(z). \quad (20)$$

The following symmetry relations are verified readily: $\theta_c(z)$ is odd when $\phi_{1x}(w)$ is even,

$$\theta_c(z) \text{ is even when } \phi_{1x}(w) \text{ is odd}; \quad (21a)$$

$\theta_s(z)$ is odd when $\phi_{1x}(w)$ is odd,

$$\theta_s(z) \text{ is even when } \phi_{1x}(w) \text{ is even}; \quad (21b)$$

The substitution of (5) in (13) leads to the following integral equation:

$$\int_{-h+\delta}^{h-\delta} \left(I_{1z}(z') \frac{e^{-j\beta_0 R_{11}}}{R_{11}} + I_{3z}(z') \frac{e^{-j\beta_0 R_{13}}}{R_{13}} \right) dz'$$

$$= \frac{-j4\pi}{\zeta_o} [C_1 \cos \beta_o z + C_2 \sin \beta_o z - \theta_A(z)]. \quad (22)$$

Since $\theta_A(z)$ is a function of $I_{2x}(x')$ and $I_{4x}(x')$, the currents in all four sides of the square are involved in (22). Three similar equations with permuted subscripts 1, 2, 3, 4 may be written. These together with (22) constitute four simultaneous integral equations in the four unknown currents. These must be evaluated in terms of the four driving voltages which are defined according to the following pattern:

$$V_{12} = \phi_1(z = -h + \delta) - \phi_2(x = -h + \delta). \quad (23)$$

Since no techniques are available for the solution of four simultaneous integral equations, it is advantageous to replace them by four independent integral equations using the method of symmetrical components. This is accomplished by expressing the driving voltages in terms of sequential combinations of four phase-sequence voltages as follows:

$$V_{12} = V^{(0)} + V^{(1)} + V^{(2)} + V^{(3)}, \quad (24a)$$

$$V_{23} = V^{(0)} + jV^{(1)} - V^{(2)} - jV^{(3)}, \quad (24b)$$

$$V_{34} = V^{(0)} - V^{(1)} + V^{(2)} - V^{(3)}, \quad (24c)$$

$$V_{41} = V^{(0)} - jV^{(1)} - V^{(2)} + jV^{(3)}. \quad (24d)$$

The four phase-sequence voltages are designated by superscripts. They may be expressed in terms of the driving voltages by solving (24a, b, c, d). The results are:

$$V^{(0)} = (1/4)(V_{12} + V_{23} + V_{34} + V_{41}), \quad (25a)$$

$$V^{(1)} = (1/4)(V_{12} - jV_{23} - V_{34} + jV_{41}), \quad (25b)$$

$$V^{(2)} = (1/4)(V_{12} - V_{23} + V_{34} - V_{41}), \quad (25c)$$

$$V^{(3)} = (1/4)(V_{12} + jV_{23} - V_{34} - jV_{41}). \quad (25d)$$

If each driving voltage is represented as in (24a, b, c, d) and the phase-sequence currents $I^{(0)}$, $I^{(1)}$, $I^{(2)}$, $I^{(3)}$ maintained by the phase-sequence voltages at each corner are determined separately, the resultant current in each side is obtained by superposition. The equations for the currents maintained by the four actual driving voltages in terms of the phase sequence currents are

$$I_1 = I^{(0)} + I^{(1)} + I^{(2)} + I^{(3)}, \quad (26a)$$

$$I_2 = I^{(0)} + jI^{(1)} - I^{(2)} - jI^{(3)}, \quad (26b)$$

$$I_3 = I^{(0)} - I^{(1)} + I^{(2)} - I^{(3)}, \quad (26c)$$

$$I_4 = I^{(0)} - jI^{(1)} - I^{(2)} + jI^{(3)}, \quad (26d)$$

or

$$I^{(0)} = (1/4)(I_1 + I_2 + I_3 + I_4), \quad (26e)$$

$$I^{(1)} = (1/4)(I_1 - jI_2 - I_3 - jI_4), \quad (26f)$$

$$I^{(2)} = (1/4)(I_1 - I_2 + I_3 - I_4), \quad (26g)$$

$$I^{(3)} = (1/4)(I_1 + jI_2 - I_3 - jI_4). \quad (26h)$$

The simultaneous determination of the four currents when arbitrary voltages are applied at the corners is thus reduced to the separate determination of four phase-sequence currents, each of which is maintained in each conductor by the appropriate phase-sequence voltage at each corner. The four independent problems are represented schematically in Fig. 8. Currents without a coordinate direction in the subscript and without an argument in parentheses have their positive direction in the clockwise sense around the square.

If one or more of the driving voltages is replaced by an impedance, the equations (24a–d) apply with the appropriate voltage replaced by the negative of the voltage drop across the impedance.

LOOP WITH FOUR EQUAL VOLTAGES IN PHASE; THE ZEROTH PHASE SEQUENCE

The zeroth phase sequence is represented by Fig. 7 when

$$V_{12} = V_{23} = V_{34} = V_{41} = V^{(0)};$$

$$I_{1z}(z) = -I_{2z}(x) = -I_{3z}(z) = I_{4z}(x) = I^{(0)};$$

$$q_1(z) = -q_2(x) = -q_3(z) = q_4(x) = q^{(0)}.$$

As a consequence of geometrical and electrical symmetry, it follows that

$$I_{1z}^{(0)}(-z) = I_{1z}^{(0)}(z); \; A_{1z}^{(0)}(-z) = A_{1z}^{(0)}(z);$$

$$A_{1z}^{(0)}(-z) = -A_{1x}^{(0)}(z); \quad (27a)$$

$$q_1^{(0)}(-z) = -q_1^{(0)}(z); \; \phi_1^{(0)}(-z) = -\phi_1^{(0)}(z), \quad (27b)$$

where the potentials are defined on the surface of conductor 1. Necessary consequences of (27) with (4) are

$$\phi_1^{(0)}(0) = 0; \; A_{1z}^{(0)}(0) = 0; \; A_{1z}^{(0)}(0) \text{ is extreme}; \quad (28a)$$

$$q_1^{(0)}(0) = 0; \; I_{1z}^{(0)}(0) \text{ is extreme}. \quad (28b)$$

It follows from (27a, b) and (21a, b) that

$$\theta_c^{(0)}(z) \text{ is even}; \; \theta_s^{(0)}(z) \text{ is odd}. \quad (29)$$

If the arbitrary lower limit in (14) is chosen to be zero; i.e., $z_o = 0$, it follows from (17) and (19) with (28a) that

$$\theta_A^{(0)}(z) = \theta_c^{(0)}(z) \cos \beta_o z + \theta_s^{(0)}(z) \sin \beta_o z \text{ is even}, \quad (30a)$$

$$\theta_V^{(0)}(z) = \theta_c^{(0)}(z) \sin \beta_o z - \theta_s^{(0)}(z) \cos \beta_o z \text{ is odd}. \quad (30b)$$

Since $A_{1z}^{(0)}(z)$ as well as $I_{1z}^{(0)}(z) = -I_{3z}^{(0)}(z)$ are even in z, it follows that both sides of the integral equation (22) must be even in z when specialized to the zeroth phase sequence. The appropriate part of the right-hand member is obtained by inspection so that the following integral equation is applicable:
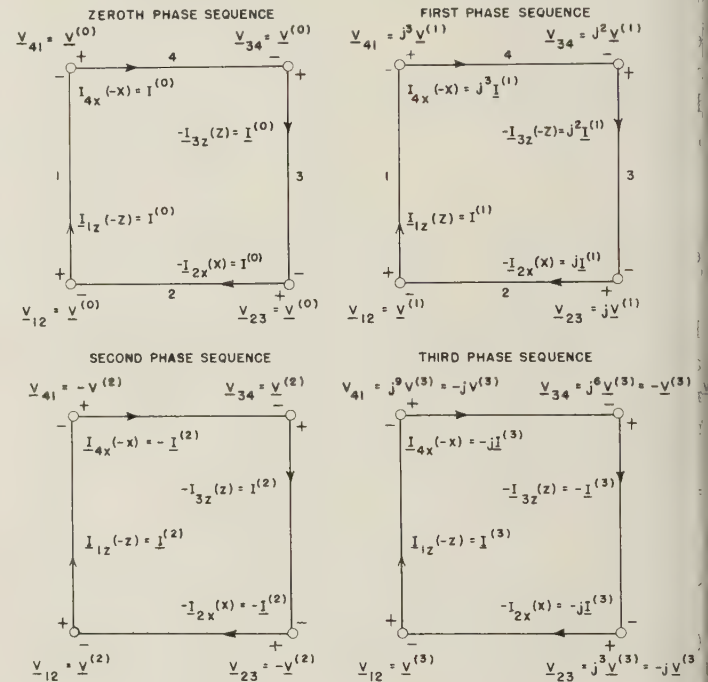
Fig. 8—The four phase sequences.

$$\int_{-h+\delta}^{h-\delta} I_{1z}^{(0)}(z') K_A^{(0)}(z, z') \, dz'$$

$$= \frac{-j4\pi}{\zeta_o} [C_1^{(0)} \cos \beta_o z - \theta_A^{(0)}(z)], \quad (31)$$

where

$$K_A^{(0)}(z, z') = \frac{e^{-j\beta_0 R_{11}}}{R_{11}} - \frac{e^{-j\beta_0 R_{12}}}{R_{13}} \quad (32a)$$

and

$$\theta_A^{(0)}(z) = \theta_s^{(0)}(z) \sin \beta_o z + \theta_c^{(0)}(z) \cos \beta_o z$$

$$= \int_o^z \phi_{1z}^{(0)}(w) \cos \beta_o(z - w)\beta_o \, dw. \quad (32b)$$

Note that $\phi_{1z}^{(0)}(z)$ may be calculated from

$$\phi_{1z}^{(0)}(z) = \frac{1}{4\pi\epsilon_o} \int_{-h+\delta}^{h-\delta} q_2^{(0)}(x') K_B^{(0)}(z, x') \, dx', \quad (33)$$

with

$$K_B^{(0)}(z, x') = \frac{e^{-j\beta_0 R_{12}}}{R_{12}} - \frac{e^{-j\beta_0 R_{14}}}{R_{14}}. \quad (34)$$

The zeroth-phase-sequence scalar potential is given by the odd part of (20). Thus,

$$\phi_1^{(0)}(z) = -C_1^{(0)} \sin \beta_o z + \theta_V^{(0)}(z), \quad (35)$$

where

$$\theta_V^{(0)}(z) = -\theta_s^{(0)}(z) \cos \beta_o z + \theta_c^{(0)}(z) \sin \beta_o z$$

$$= \int_o^z \phi_{1x}^{(0)}(w) \sin \beta_o(z - w)\beta_o \, dw. \quad (36)$$

The driving voltage is obtained from (20) and (23) with (27b). Note that $\phi_2^{(0)}(x) = -\phi_1^{(0)}(z)$, so that

$$\tfrac{1}{2} V^{(0)} = \phi_1^{(0)}(z = -h + \delta) =$$
$$-C_1^{(0)} \sin \beta_o(-h + \delta) + \theta_V^{(0)}(-h + \delta). \quad (37)$$

The integral equation (31) for the zeroth-phase-sequence current may be solved by iteration in a manner similar to that used for the center-driven dipole.[11,12] In order to accomplish this let

$$\Psi^{(0)}(z) = \Psi^{(0)} + \gamma^{(0)}(z)$$

$$= \int_{-h+\delta}^{h-\delta} g^{(0)}(z, z') K_A^{(0)}(z, z') \, dz', \quad (38)$$

where

$$g^{(0)}(z, z') = I_z^{(0)}(z')/I_z^{(0)}(z), \quad (39)$$

and $\Psi^{(0)}$ is the sensibly constant magnitude of $\Psi^{(0)}(z)$ at an appropriately chosen point—preferably the point of maximum vector potential. As a consequence of the symmetry conditions (27a, b) and (28), the maximum of the axial component of the vector potential is at $z = 0$. Therefore, $\Psi^{(0)}$ may be defined at once. It is

$$\Psi^{(0)} = \left| \Psi^{(0)}(z = 0) \right|. \quad (40)$$

The function $\gamma^{(0)}(z)$ is the difference between $\Psi^{(0)}(z)$ and the magnitude $\left| \Psi^{(0)}(z = 0) \right|$. The distribution function $g^{(0)}(z, z')$ is the ratio of the current at any point z' to the current at any other point z.

By adding $I_z^{(0)}\Psi^{(0)}(z)$ to both sides of (31) and rearranging terms, the following equivalent equation is obtained

$$I_z^{(0)} = \frac{-j4\pi}{\zeta_o \Psi^{(0)}} \left[C_1^{(0)} \cos \beta_o z - \theta_A^{(0)}(z) \right]$$

$$- \frac{1}{\Psi^{(o)}} \left\{ \int_{-h+\delta}^{h-\delta} \left[I_z^{(0)}(z') - I_z^{(0)}(z)g^{(0)}(z, z') \right] \cdot \right.$$

$$\left. \times K_A^{(0)}(z, z') \, dz' + I_z^{(0)}(z)\gamma^{(0)}(z) \right\}. \quad (41)$$

If $g^{(0)}(z, z')$ is a good approximation of the current distribution, the brace on the right is small. The term $\theta_A^{(0)}(z)$ as given in (15) depends upon the contribution to the scalar potential by the charges on the adjacent sides of the square. These are significant primarily near the corners where the sides are close together. Therefore, it is reasonable to assume that $\theta_A^{(0)}(z)$ is not a major factor in the over-all distribution of current and may be omitted in a zeroth-order approximation. It follows that satisfactory zeroth-order distributions of current and

[11] R. King and D. Middleton, "Cylindrical antenna: current and impedance," *Quart. Appl. Math.*, vol. 3, pp. 302–335; 1946, vol. 4, pp. 199; 1946.

[12] R. W. P. King, "Theory of Linear Antennas, Harvard University Press, Cambridge, Mass., Chap. II; 1956.

charge and the associated potentials are

$$[I_z^{(0)}(z)]_o = \frac{-j4\pi}{\zeta^0 \Psi^{(0)}} C_1^{(0)} F_o(z);$$

$$[q^{(0)}(z)]_o = \frac{-4\pi\epsilon_o}{\Psi^{(0)}} C_1^{(0)} G_o(z); \quad (42a)$$

$$[A_z^{(0)}(z)]_o = \frac{-j}{v_o} C_1^{(0)} F_o(z);$$

$$[\phi^{(0)}(z)]_o = [\phi_z^{(0)}(z)]_o = C_1^{(0)} G_o(z). \quad (42b)$$

The shorthand

$$F_o(z) = \cos \beta_o z, \quad G_o(z) = \sin \beta_o z, \quad (43)$$

is used. Zeroth-order distributions of current in the zeroth-phase sequence are shown in Fig. 9.
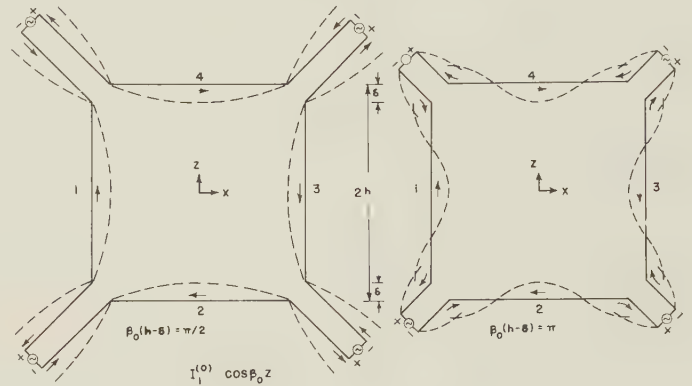
Fig. 9—Zeroth phase-sequence currents on a square loop.

A first-order solution is obtained by substituting the zeroth-order current and charge distributions from (42a) into the integrals in (41). With (33) and (42), the following first-order expressions may be derived. Note that $q_2(x') = -q_1(z')$.

$$[\phi_{1x}^{(0)}(z)]_1 = -\frac{C_1^{(0)}}{\Psi^{(0)}} J_1^{(0)}(z),$$

$$J_1^{(0)}(z) = -\int_{-h+\delta}^{h-\delta} \sin \beta_o x' K_B^{(0)}(z, x') \, dx'; \quad (44a)$$

$$[\theta_A^{(0)}(z)]_1 = -\frac{C_1^{(0)}}{\Psi^{(o)}} L_1^{(0)}(z),$$

$$L_1^{(0)}(z) = \int_o^z J_1^{(0)}(w) \cos \beta_o(z - w)\beta_o \, dw. \quad (44b)$$

With (41) and (44a, b), the following first-order solution for the current in the zeroth-phase sequence is obtained

$$[I_{1z}^{(0)}(z)]_1$$

$$= \frac{-j4\pi}{\zeta_o \Psi^{(0)}} C_1^{(0)} \left\{ F_o(z) + \frac{1}{\Psi^{(0)}} \left[F_1^{(o)}(z) + L_1^{(0)}(z) \right] \right\}, \quad (45)$$

where

$$F_1^{(0)}(z)$$

$$= -\left\{ \int_{-h+\delta}^{h-\delta} F_o(z') - F_o(z)g^{(0)}(z, z')] K_A^{(0)}(z, z') \, dz' \right.$$

$$\left. + F_o(z)\gamma^{(0)}(z) \right\}. \quad (46)$$

A first-order evaluation of the constant $C_1^{(0)}$ may be achieved using (36) and (37). Thus,

$$\tfrac{1}{2} V^{(0)}$$
$$= -[C_1^{(0)}]_1 \left[\sin \beta_o(-h + \delta) + \frac{M_1^{(0)}(-h + \delta)}{\Psi^{(0)}} \right], \quad (47)$$

where

$$M_i^{(0)}(-h + \delta)$$
$$= \int_o^{-h+\delta} J_1^{(0)}(w) \sin \beta_o(-h + \delta - w)\beta_o \, dw. \quad (48)$$

By solving (47) for the first-order value $[C_1^{(0)}]_1$ and substituting this in (45), the following first-order formula for the zeroth phase-sequence current is obtained:

$$[I_{1z}^{(0)}(z)]_1$$
$$= \frac{j2\pi V^{(0)}}{\zeta_o \Psi^{(0)}} \left\{ \frac{\cos \beta_o z + [F_1^{(0)}(z) + L_1^{(0)}(z)]/\Psi^{(0)}}{\sin \beta_o(-h+\delta) + M_1^{(0)}(-h+\delta)/\Psi^{(0)}} \right\} \quad (49)$$

Second- and still higher-order terms may be derived by continuing the iteration. Note that this also involves the determination of second-and higher-order values of the constant $C_1^{(0)}$.

The expansion parameter for the zeroth-phase sequence may be obtained by using the zeroth-order current from (42a) or (49) in (39). That is, let

$$g^{(0)}(z, z') = \cos \beta_o z'/\cos \beta_o z, \quad (50)$$

so that (38) becomes

$$\Psi^{(0)}(z) = \Psi_1^{(0)} + \gamma_1^{(0)}(z)$$
$$= \sec \beta_o z \int_{-h+\delta}^{h-\delta} \cos \beta_o z' K_A^{(0)}(z, z') \, dz', \quad (51)$$

where $K_A^{(0)}(z, z')$ is defined in (32a). In evaluating $\Psi_1^{(0)}(z)$ the small distance δ may be neglected and the integration extended from $-h$ to h. The result is

$$\Psi_1^{(0)}(z) \doteq \sec \beta_o z[C_a(h, z) - C_{2h}(h, z)], \quad (52a)$$

where

$$C_i(h, z) = \int_{-h}^{h} \cos \beta_o z' \frac{e^{-j\beta_o R_i}}{R_i} \, dz';$$

$$R_i = \sqrt{(z - z')^2 + i^2}; \quad (52b)$$

with $i = a, 2h$. The function $C_i(h, z)$ has been evaluated[11,12] in terms of the tabulated generalized sine and cosine integrals.[13] It is predominantly real if h/a is large, as is assumed, and its magnitude is sensibly constant at its value when $z = o$. Thus, let $\Psi_1^{(0)}$ be defined as follows:

$$\Psi_1^{(0)} = |\Psi_1^{(0)}(z = 0)| = |C_a(h, 0) - C_{2h}(h, 0)|. \quad (52c)$$

Since $2h$ is great compared with a, the second term in (52c) is relatively small compared with the first term.

It is readily verified that for an electrically small loop that satisfies the condition, $\beta_o^2 h^2 << 1$, the following values are good approximations:

13 "Tables of Generalized Sine- and Cosine-Integral Functions," Harvard University Press, Cambridge, Mass., vols. I and II; 1949.

$$C_a(h, 0) \doteq 2ln \frac{2h}{a} = \Omega;$$

$$C_{2h}(h, 0) \doteq ln \left[\frac{\sqrt{5} + 1}{\sqrt{5} - 1} \right] \doteq 0.96.$$

Hence,

$$\Psi_1^{(o)} \doteq \Omega - 0.96 \doteq \Omega - 1. \quad (53)$$

Since the same zeroth-order distribution of current is used in $g^{(0)}(z, z')$ as in the evaluation of the first-order integrals, the difference integral in (46) vanishes so that $F_1^{(0)}(z) = -F_o(z)\gamma^{(0)}(z)$. The correction factor $\gamma^{(0)}(z)$ is very small except where the current vanishes. This may occur along the loop with the zeroth-order distribution of current used in defining $\Psi^{(0)}$ and $\gamma^{(0)}(z)$, but not with a higher-order approximation. As shown in the literature for the dipole antenna,[11,12] the use of a first-order current instead of a zeroth-order value eliminates the apparent infinities of $\Psi(z)$ occurring at the driving point for certain values of $\beta_o h$ and reveals an essentially constant behavior with $|\Psi(z)| \doteq \Psi$ for all values of h, where Ψ is the value obtained using the zeroth-order current at its maximum value. The same situation obtains in the case of the loop, and the apparent infinities in $\gamma^{(0)}(z)$ are a consequence of the inadequacy of the zeroth-order currents where these vanish or become very small. Actually, the term $F_o(z)\gamma^{(0)}(z)$ is always small, and with it the first-order value $F_1(z)$. The first-order current is

$$[I_{1z}^{(0)}(z)]_1$$
$$= \frac{j2\pi V^{(0)}}{\zeta_o \Psi_1^{(0)}} \left\{ \frac{\cos \beta_o z + N_1^{(0)}(z) - \gamma^{(0)}(z) \cos \beta_o z]/\Psi_1^{(0)}}{\sin \beta_o(-h + \delta) + M_1^{(0)}(-h + \delta)/\Psi_1^{(0)}} \right\}. \quad (54)$$

The first-order input admittance is obtained by setting $z = -h + \delta$. Thus

$$[Y_{in}^{(0)}]_1 = \frac{j2\pi}{\zeta_o \Psi_1^{(o)}}$$
$$\times \left\{ \frac{\cos \beta_o(-h + \delta) + N_1^{(0)}(-h + \delta)/\Psi_1^{(0)}}{\sin \beta_o(-h + \delta) + M_1^{(0)}(-h + \delta)/\Psi_1^{(0)}} \right\} \quad (55a)$$

where

$$N_1^{(0)} = L_1^{(0)}(-h + \delta) - \gamma^{(0)}(-h + \delta) \cos \beta_o(-h + \delta).$$

The functions $L_1^{(0)}(-h + \delta)$, $M_1^{(0)}(-h + \delta)$ may be expressed in terms of arc hyperbolic functions and simple integrals of the tabulated generalized sine- and cosine-integral functions.[13] These may be evaluated by numerical methods. This completes the formal solution to a first-order of the zeroth phase sequence. Note that the zeroth-order input impedance is

$$[Z_{in}^{(0)}]_o = \frac{j\zeta_o \Psi^{(0)}}{2\pi} \tan \beta_o(h - \delta). \quad (55b)$$

LOOP WITH FOUR EQUAL VOLTAGES ALTERNATELY OPPOSITE IN PHASE: THE SECOND PHASE SEQUENCE

The second phase sequence—in which the four corner voltages are equal in magnitude but alternately opposite in phase—is more closely related to the zeroth

nase sequence than are the first and fourth phase
quences. It is represented in Fig. 7 when

$$V_{12} = -V_{23} = V_{34} = -V_{41} = V^{(2)};$$

$$I_{1z}(z) = -I_{2x}(x) = I_{3z}(z) = -I_{4x}(x) = I^{(2)};$$

$$q_1(z) = -q_2(x) = q_3(z) = -q_4(x) = q^{(2)}.$$

s a consequence of electrical and geometrical symmetry,
follows that

$$_{1z}^{(2)}(-z) = -I_{1z}^{(2)}(z); \quad A_{1z}^{(2)}(-z) = -A_{1z}^{(2)}(z);$$

$$A_{1x}^{(2)}(-z) = A_{1x}^{(2)}(z); \quad (56a)$$

$$q_1^{(2)}(-z) = q_1^{(2)}(z); \quad \phi_1^{(2)}(-z) = \phi_1^{(2)}(z); \quad (56b)$$

here the potentials are defined on the surface of con-
uctor 1. Necessary consequences of (56a, b) and (2b)
re

$$A_{1z}^{(2)}(0) = 0; \quad \phi_1^{(2)}(0) \text{ is extreme}; \quad (57a)$$

$$I_{1z}^{(2)}(0) = 0; \quad q_1^{(2)}(0) \text{ is extreme}. \quad (57b)$$

t follows from (56a, b) and (21a, b) that

$$\theta_c^{(2)}(z) \text{ is odd}; \quad \theta_s^{(2)}(z) \text{ is even}; \quad (58a)$$

o that with the arbitrary lower limit z_o in (14) again
hosen to be zero, it is evident from (17) and (19) that

$$\theta_A^{(2)}(z) \text{ is odd}; \quad \theta_V^{(2)}(z) \text{ is even}. \quad (58b)$$

ince $A_{1z}^{(2)}(z) = -A_{1z}^{(2)}(-z)$, the odd part of the right
ide of (22) must be selected. This is obtained by in-
pection. Thus, the integral equation for the current in
he second phase sequence is

$$\int_{-h+\delta}^{h-\delta} I_{1z}^{(2)}(z') K_A^{(2)}(z, z') \, dz'$$

$$= \frac{-j4\pi}{\zeta_o} \left[C_2^{(2)} \sin \beta_o z - \theta_A^{(2)}(z) \right], \quad (59a)$$

vhere

$$K_A^{(2)}(z, z') = \frac{e^{-j\beta_0 R_{11}}}{R_{11}} + \frac{e^{-j\beta_0 R_{13}}}{R_{13}} : \quad (59b)$$

$$_A^{(2)}(z) = \left[\theta_s^{(2)}(z) - \phi_{1x}^{(2)}(0) \right] \sin \beta_o z$$

$$+ \theta_c^{(2)}(z) \cos \beta_o z \quad (60a)$$

$$= \int_0^z \phi_{1x}^{(2)}(w) \cos \beta_o(z - w)\beta_o \, dw$$

$$- \phi_{1x}^{(2)}(0) \sin \beta_o z. \quad (60b)$$

In (60b) $\phi_{1x}^{(2)}(w)$ may be calculated from

$$\phi_{1x}^{(2)}(z) = \frac{1}{4\pi\epsilon_o} \int_{-h+\delta}^{h-\delta} q_2^{(2)}(x') K_B^{(2)}(z, x') \, dx' \quad (61)$$

with

$$K_B^{(2)}(z, x') = \frac{e^{-j\beta_0 R_{12}}}{R_{12}} + \frac{e^{-j\beta_0 R_{14}}}{R_{14}}. \quad (62)$$

The expressions for the scalar potential and the
second phase sequence voltage are obtained from (20)
and (23). Thus,

$$\tfrac{1}{2} V^{(2)} = \phi_1^{(2)}(z = -h + \delta)$$

$$= C_2^{(2)} \cos \beta_o(-h + \delta) + \theta_V^{(2)}(-h + \delta), \quad (63)$$

where $\theta_V^{(2)}(-h + \delta)$ is obtained by setting $z = -h + \delta$
in the following expression:

$$\theta_V^{(2)}(z) = \int_0^z \phi_{1x}^{(2)}(w) \sin \beta_o(z - w)\beta_o \, dw$$

$$+ \phi_{1x}^{(2)}(0) \cos \beta_o z. \quad (64)$$

The integral equation (59a) for the second phase-
sequence current may be solved by iteration following
the procedure used for (31). Thus, let

$$\Psi^{(2)}(z) = \Psi^{(2)} + \gamma^{(2)}(z)$$

$$= \int_{-h+\delta}^{h-\delta} g^{(2)}(z, z') K_A^{(2)}(z, z') \, dz' \quad (65)$$

where

$$g^{(2)}(z, z') = I_z^{(2)}(z')/I_z^{(2)}(z), \quad (66)$$

and where $\Psi^{(2)}$ is the sensibly constant magnitude of
$\Psi^{(2)}(z)$. For the expansion parameter $\Psi_1^{(0)}$ it was pos-
sible to select the value of $\Psi^{(0)}(z)$ at $z = 0$ where vector
potential and current are both extreme. Although the
vector potential and the current for the second phase-
sequence distribution both vanish at $z = 0$, symmetry
is such that their ratio is essentially constant and may
be used in the definition of $\Psi^{(2)}$.

By adding $I_z^{(2)} z \Psi^{(2)}(z)$ to both sides of (59) and rear-
ranging, the following equation is obtained:

$$I_z^{(2)} z = \frac{-j4\pi}{\zeta_o \Psi^{(2)}} \left[C_2^{(2)} \sin \beta_o z - \theta_A^{(2)}(z) \right]$$

$$- \frac{1}{\Psi^{(2)}} \left[\int_{-h+\delta}^{h-\delta} \left[I_z'^{(2)} - I_z z g^{(2)}(z, z') \right] \right.$$

$$\left. K_A^{(2)}(z, z') \, dz' + I_z^{(2)} \gamma^{(2)}(z) \right]. \quad (67)$$

As in the case of the zero phase sequence, the leading
term in the distribution of current is the triginometric
one in (67). Thus, satisfactory zeroth-order distributions
of current and charge are

$$\left[I_z^{(2)}(z) \right]_o = \frac{-j4\pi}{\zeta_o \Psi^{(2)}} C_2^{(2)} G_o(z);$$

$$\left[q^{(2)}(z) \right]_o = \frac{4\pi\epsilon_o}{\Psi^{(2)}} C_2^{(2)} F_o(z). \quad (68)$$
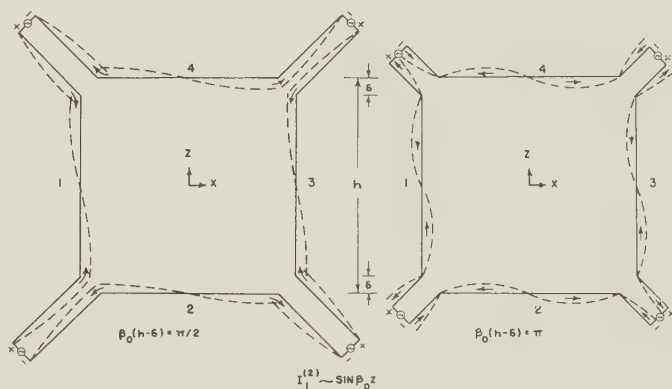
This distribution is illustrated in Fig. 10.

Fig. 10—Second phase-sequence currents on a square loop.

The associated zeroth-order potentials on the surfaces
of the conductor are

$$\left[A_{1z}^{(2)}(z) \right]_o = \frac{-j}{v_o} C_2^{(2)} G_o(z); \quad \left[\phi_1^{(2)}(z) \right]_o = C_2^{(2)} F_o(z). \quad (69)$$

A first-order solution may be obtained by substituting
(68) into the integrals in (67). With (61) the following

first-order expressions are obtained[14]:

$$\left[\phi_{1x}^{(2)}(z)\right]_1 = \frac{C_2^{(2)}}{\Psi^{(2)}} J_1^{(2)}(z);$$

$$J_1^{(2)}(z) = -\int_{-h+\delta}^{h-\delta} \cos \beta_o x' K_B^{(2)}(z, x')\, dx'; \quad (70a)$$

$$\left[\theta_A^{(2)}(z)\right]_1 = -\frac{C_2^{(2)}}{\Psi^{(2)}} L_1^{(2)}(z);$$

$$L_1^{(2)}(z) = -\int_o^z J_1^{(2)}(w) \cos \beta_o(z - w)\beta_o\, dw$$
$$+ J_1^{(2)}(0) \sin \beta_o z. \quad (70b)$$

With (67) and (70a, b) the following first-order solution for the current in the second phase sequence is obtained:

$$\left[I_{1z}^{(2)}(z)\right]_1 = \frac{-j4\pi}{\zeta_o \Psi^{(2)}} C_2^{(2)} \Big\{ G_o(z)$$
$$+ \frac{1}{\Psi^{(2)}} \left[G_1^{(2)}(z) + L_1^{(2)}(z)\right] \Big\}, \quad (71)$$

where

$$G_1^{(2)}(z) = -\Big\{ \int_{-h+\delta}^{h-\delta} \left[G_o(z')\right.$$
$$- G_o(z)g^{(2)}(z, z')\right] K_A^{(2)}(z, z')\, dz' + G_o(z)\gamma^{(2)}(z) \Big\}. \quad (72)$$

A first-order determination of the constant $C_2^{(2)}$ is achieved using (63) and (64). Thus,

$$\tfrac{1}{2} V^{(2)} = \left[C_2^{(2)}\right]_1 \left[\cos \beta_o(-h + \delta) + \frac{M_1^{(2)}(-h + \delta)}{\Psi^{(2)}} \right], \quad (73)$$

where

$$M_1^{(2)}(-h + \delta) = \int_o^{-h+\delta} J_1^{(2)}(w)$$
$$\sin \beta_o(-h + \delta - w)\beta_o\, dw + J_1^{(2)}(0) \cos \beta_o(-h + \delta). \quad (74)$$

The first-order constant $\left[C_2^{(2)}\right]_1$ is determined from (73). It may be substituted in (71) to give:

$$\left[I_{1z}^{(2)}(z)\right]_1$$
$$= \frac{-j2\pi V^{(2)}}{\zeta_o \Psi^{(2)}} \Big\{ \frac{\sin \beta_o z + \left[G_1^{(2)}(z) + L_1^{(2)}(z)\right]/\Psi^{(2)}}{\cos \beta_o(-h + \delta) + M_1^{(2)}(-h + \delta)/\Psi^{(2)}} \Big\}. \quad (75)$$

Second- and higher-order terms may be obtained by continuing the iteration. This also involves higher-order values of $C_2^{(2)}$.

The expansion parameter for the second phase sequence may be obtained by using the zeroth-order current from (68) in (65). Thus with

$$g^{(2)}(z, z') = \sin \beta_o z'/\sin \beta_o z \quad (76)$$

it follows that

$$\Psi_1^{(2)}(z) = \Psi_1^{(2)} + \gamma_1^{(2)}(z)$$
$$= \left[S_a(h, z) + S_{2h}(h, z)\right] \csc \beta_o z, \quad (77)$$

where

$$S_i(h, z) = \int_{-h}^h \sin \beta_o z' \frac{e^{-j\beta_o R_1}}{R_i}; \quad R_i = \sqrt{(z - z')^2 + i^2}$$
$$(78)$$

<hr>

[14] Note that $q_2(x) = -q_1(z)$.

with $i = a, 2h$. The small quantity δ has been neglected. The sensibly constant magnitude $\Psi_1^{(2)}$ of $\Psi_1^{(2)}(z)$ may be defined as follows

$$\Psi_1^{(2)} = \lim_{z \to o} \left| \Psi(z) \right|. \quad (79)$$

Since the numerator and denominator in (79) both vanish at $z = 0$, each may be differentiated with respect to z in order to evaluate the indeterminate form.

$$\Psi_1^{(2)} = \left| C_a(h, 0) + C_{2h}(h, 0) \right.$$
$$\left. - \frac{2 \sin \beta_o h}{\beta_o h} K_A^{(2)}(0, h) \right|. \quad (80)$$

This does not differ greatly from $\Psi_1^{(o)}$. For the electrically small loop ($\beta_o h \ll 1$), the formulas preceding (53) may be used in (80). The result is

$$\Psi_1^{(2)} \doteq \Omega - 1.04 \doteq \Omega - 1 \doteq \Psi_1^{(0)}. \quad (81)$$

Since the same zeroth-order distribution of current is used in $g^{(2)}(z, z')$ as in the evaluation of the first-order integrals, the difference integral in (72) vanishes to that $G_1^{(2)}(z) = -G_o(z)\gamma^{(2)}(z)$. This is a very small term. The first-order current and admittance for the second-phase sequence are

$$\left[I_{1z}^{(2)}(z)\right]_1 = \frac{-j2\pi V^{(2)}}{\zeta_o \Psi_1^{(2)}}$$
$$\times \Big\{ \frac{\sin \beta_o z + \left[L_1^{(2)}(z) - \gamma^{(2)}(z) \sin \beta_o z\right]/\Psi_1^{(2)}}{\cos \beta_o(-h + \delta) + M_1^{(2)}(-h + \delta)/\Psi_1^{(2)}} \Big\} \quad (82)$$

$$\left[Y_{in}^{(2)}\right]_1 = \frac{-j2\pi}{\zeta_o \Psi_1^{(2)}}$$
$$\times \Big\{ \frac{\sin \beta_o(-h + \delta) + N_1^{(2)}(-h + \delta)/\Psi_1^{(2)}}{\cos \beta_o(-h + \delta) + M_1^{(2)}(-h + \delta)/\Psi_1^{(2)}} \Big\} \quad (83)$$

where

$$N_1^{(2)}(-h + \delta) = L_1^{(2)}(-h + \delta) \sin \beta_o(-h + \delta)$$

The functions $L_1^{(2)}(-h + \delta)$ and $M_1^{(2)}(-h + \delta)$ are similar to the corresponding functions for the zeroth phase sequence and may be evaluated in the same manner. This completes the formal solution to a first order of the second phase sequence. Note that the zeroth-order impedance is

$$\left[Z_{in}^{(2)}\right]_o = \frac{-j\zeta_o \Psi_1^{(2)}}{2\pi} \cot \beta_o(h - \delta). \quad (84)$$

SUPERPOSITION OF THE ZEROTH AND SECOND PHASE SEQUENCES

A number of interesting special cases of the corner driven square loop antenna may be treated in terms of a superposition of the zeroth and second phase sequences only.

If the first and third phase sequences are not excited so that $V^{(1)} = V^{(3)} = 0$, it follows from (24a-d) that

$$V_{12} = V_{34} = V^{(0)} + V^{(2)};$$
$$V_{23} = V_{41} = V^{(0)} - V^{(2)}; \quad (85a)$$

and from (26a),

$$I_1 = I_3 = I^{(0)} + I^{(2)}; \quad I_2 = I_4 = I^{(0)} - I^{(2)}. \quad (85b)$$

In terms of arbitrary voltages $V_{12} = V_{34}$ and $V_{23} = V_{41}$, the phase-sequence voltages are

$$V^{(0)} = \tfrac{1}{2}(V_{12} + V_{23}); \qquad V^{(2)} = \tfrac{1}{2}(V_{12} - V_{23}). \quad (85c)$$

Since at each corner $I^{(0)} = Y^{(0)} V^{(0)}$, $I^{(2)} = Y^{(2)} V^{(2)}$, it follows with (85b) that at corners 12 and 23 the currents are

$$I_{1z}(-h + \delta) = I_1 = V_{12} Y_s + V_{23} Y_t ; \quad (86a)$$

$$-I_{2x}(-h + \delta) = I_2 = V_{12} Y^t + V_{23} Y_s ; \quad (86b)$$

where

$$Y_s = \tfrac{1}{2}(Y^{(0)} + Y^{(2)}); \quad Y_t = \tfrac{1}{2}(Y^{(0)} - Y^{(2)}). \quad (86c)$$

If a load or tuning admittance $Y_2 = 1/Z_2$ is connected in place of the driving voltages V_{23} and V_{41}, it is possible according to the Compensation Theorem to set $V_{23} = -I_2 Z_2$ or $I_2 = -Y_2 V_{23}$ in (86b) to obtain

$$I_{1z}(-h + \delta) = I_1 = V_{12}\Big[Y_s - \frac{Y_t^2}{Y_s + Y_2}\Big] = V_{12} Y_{in};$$

$$-I_{2z}(-h + \delta) = I_2 = V_{12}\frac{Y_t Y_2}{Y_s + Y_2}. \quad (86d)$$

$$I_{4x}(X) = I^{(0)}(x) - I^{(2)}(-x)$$
$$-I_{3z}(z) = I^{(0)}(z) + I^{(2)}(z)$$
$$I_{1z}(-z) = I^{(0)}(-z) + I^{(2)}(-z)$$
$$-I_{2x}(x) = I^{(0)}(x) - I^{(2)}(x)$$

(a)

(b) $\beta_o h = 0.6\pi$

GROUND SCREEN

(c) $\beta_o h = 0.6\pi$
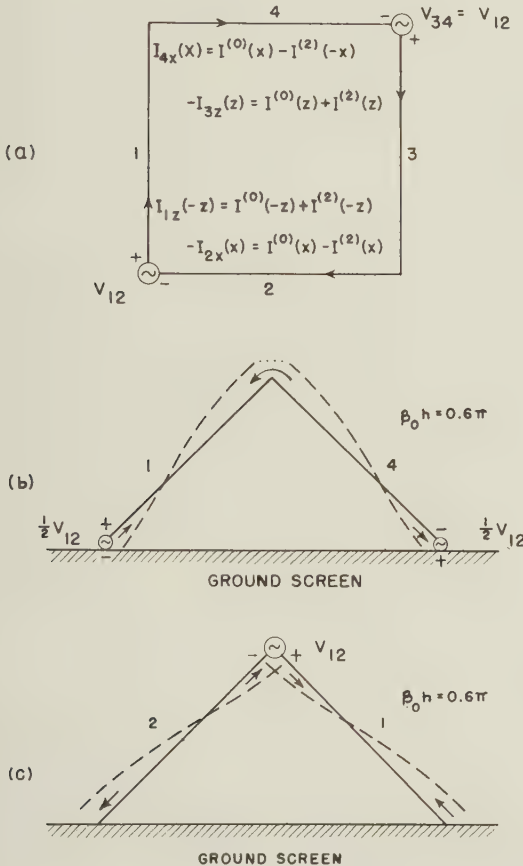
GROUND SCREEN

Fig. 11—Square loop antenna driven in phase at one pair of diagonal corners.

The simplest case is obtained when $V^{(0)} = V^{(2)}$ so that $V_{23} = V_{41} = 0$; $V_{12} = V_{34} = 2V^{(0)}$. These conditions correspond to a square loop driven by equal voltages in phase at two corners at opposite ends of one diagonal. This is illustrated in Fig. 11(a) and also in Figs. 11(b), (c) where two possible arrangements of

half of a square loop over an image plane are shown. Another possibility is the corner-reflector loop consisting of one side of a square loop over two mutually perpendicular image planes as shown in Fig. 12.

The first-order current is obtained by combining (75) and (54) to give

$$[I_{1z}(z)]_1 = [I_{1z}^{(0)}(z) + I_{1z}^{(2)}(z)]_1 . \quad (86e)$$

Since no significant simplification is possible in this first-order formula it is not written out. On the other
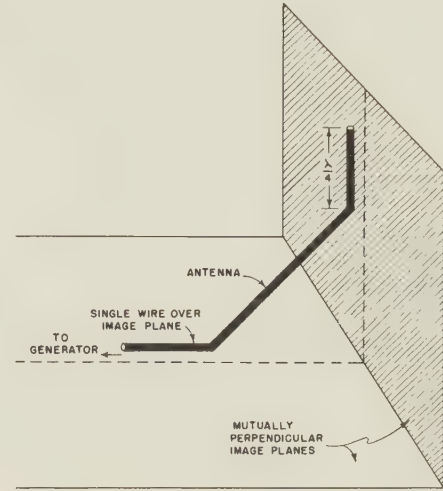

Fig. 12—Corner reflector loop.

hand, the zeroth-order current has a simple approximate form. Thus

$$[I_{1z}(z)]_o = \frac{j\pi V_{12}}{\zeta_o \Psi_1^{(0)} \Psi_1^{(2)}} \quad (87a)$$

$$\times \left[\frac{\Psi_1^{(2)} \cos \beta_o(-h + \delta) \cos \beta_o z - \Psi_1^{(0)} \sin \beta_o(-h + \delta) \sin \beta_o z}{\tfrac{1}{2} \sin 2\beta_o(-h + \delta)}\right].$$

In general, for thin conductors, $\Psi_1^{(2)}$ and $\Psi_1^{(0)}$ do not differ greatly. Let Ψ and Δ be defined as follows:

$$\frac{2}{\Psi} = \left[\frac{1}{\Psi_1^{(0)}} + \frac{1}{\Psi_1^{(2)}}\right]; \quad \frac{2\Delta}{\Psi} = \left[\frac{1}{\Psi_1^{(2)}} - \frac{1}{\Psi_1^{(0)}}\right];$$

$$\frac{1 + \Delta}{\Psi} = \frac{1}{\Psi_1^{(2)}} ; \quad \frac{1 - \Delta}{\Psi} = \frac{1}{\Psi_1^{(0)}} ; \quad (87b)$$

so that

$$[I_{1z}(z)]_o = \frac{j2\pi V_{12}}{\zeta_o \Psi}$$

$$\times \left[\frac{\cos \beta_o(-h + \delta + z) + \Delta \cos \beta_o(-h + \delta - z)}{\sin 2\beta_o(-h + \delta)}\right] \quad (87c)$$

Since Δ is small, the zeroth-order distribution of current is approximately that given by the first term in the square brackets. This is illustrated in Figs. 11b, c, and also in Figs. 13 and 14.

The input admittance is given by $I_{1z}(-h + \delta)/V_{12}$. The first-order value is

$$[Y_{in}]_1 = \tfrac{1}{2}\{[Y_{in}^{(0)}] + [Y_{in}^{(2)}]_1\}, \quad (87d)$$

where $[Y_{in}^{(0)}]_1$ is given by (55) and $[Y_{in}^{(2)}]_1$ is in (83). The approximate zeroth-order formula for the input impedance as obtained from (87c) by neglecting the small term with Δ as a factor is:

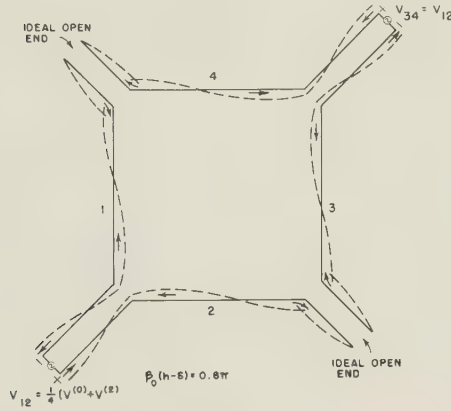$$[Z_{in}]_1 = \frac{j\zeta_o \Psi}{2\pi} \tan 2\beta_o(h - \delta). \qquad (87e)$$

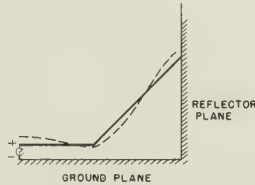Fig. 13—Superposition of zeroth and second phase-sequence currents in a square loop.

Fig. 14—One member of square loop in corner reflector.

FIRST AND THIRD-PHASE SEQUENCES—DIAGONAL DIPOLE MODES

It is seen from Fig. 8 that both the first- and the third-phase sequences may be treated as superpositions of the two circuits shown in Fig. 15. A superposition of the two currents obtained from Figs. 15(a) and 15(b) yields the currents of the first phase-sequence directly. Currents in the third phase-sequence are obtained in the same manner after reversing the sign of V in Fig. 15(b). For convenience let the distribution obtained from the arrangement of Fig. 15(a) or Fig. 15(b) be called the *diagonal dipole mode of the loop*. In terms of the notation of Fig. 15 the currents of the first- and third-phase sequences in the conductor 1 are, respectively,

$$I_{1z}^{(1)}(z) = I_{1z}^{(a)}(z) + I_{1z}^{(b)}(z), \qquad (88a)$$

$$I_{1z}^{(3)}(z) = I_{1z}^{(a)}(z) - I_{1z}^{(b)}(z), \qquad (88b)$$

where $I_{1z}^{(a)}(z)$ is the current in conductor 1 in the circuit in Fig. 15a, $I_{1z}^{(b)}(z)$ the current in conductor 1 in Fig. 15(b). Note that

$$I_{1z}^{(b)}(z) = jI_{4x}^{(a)}(x), \qquad (88c)$$

and from symmetry,

$$I_{4x}^{(a)}(x) = -I_{1z}^{(a)}(-z). \qquad (88d)$$

Hence,

$$I_{1z}^{(1)}(z) = I_{1z}^{(a)}(z) - jI_{1z}^{(a)}(-z) \qquad (88e)$$

$$I_{1z}^{(3)}(z) = I_{1z}^{(a)}(z) + jI_{1z}^{(a)}(-z). \qquad (88f)$$

Thus, it is sufficient to determine $I_{1z}^{(a)}(z)$ in Fig. 5a for the entire range of z between $-h + \delta$ and $h - \delta$.

The current $I_{1z}^{(a)}(z)$ in Fig. 15(a) may be determined following essentially the same procedure used in solving for the current in a V-antenna [7]. However, the two V-antennas are coupled. Each diagonal half of the square in Fig. 15(a) is a right-angled V-antenna with vanishing currents at the ends. That is,

$$I_{1z}^{(a)}(h - \delta) = I_{2x}^{(a)}(h - \delta)$$
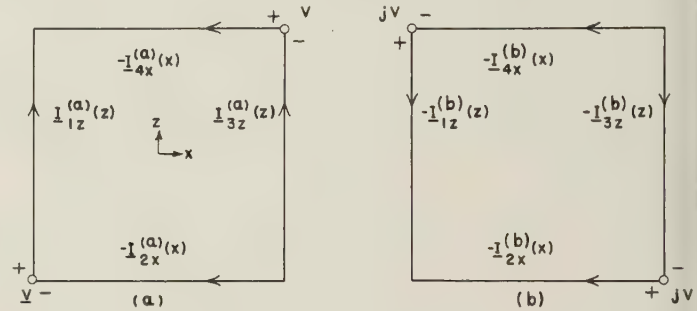$$= I_{3z}^{(a)}(-h + \delta) = I_{4x}^{(a)}(-h + \delta) = 0. \qquad (89a)$$

Fig. 15—Components of the first- and third-phase sequences.

The following symmetry conditions obtain:

$$I_{1z}^{(a)}(z) = -I_{2x}^{(a)}(x) = I_{3z}^{(a)}(-z) = -I_{4x}^{(a)}(-x); \qquad (89b)$$

$$q_1^{(a)}(z) = -q_2^{(a)}(x) = -q_3^{(a)}(-z) = q_4^{(a)}(-x). \qquad (89c)$$

The components of the vector potential on the surfaces of the conductors satisfy the same symmetry conditions as the currents; the scalar potentials satisfy the same conditions as the charges per unit length.

The equation for the vector potential and integral equation for the current is (22) with (89b). That is

$$4\pi \nu_o A_{1z}(z) = \int_{-h+\delta}^{h-\delta} \left[I_{1z}(z') \frac{e^{-j\beta_0 R_{11}}}{R_{11}} \right.$$

$$\left. + I_{1z}(-z') \frac{e^{-j\beta_0 R_{13}}}{R_{13}} \right] d\dot{z}'$$

$$= \frac{-j4\pi}{\zeta_o} \left[C_1 \cos \beta_o z + C_2 \sin \beta_o z - \theta_A(z) \right], \qquad (90a)$$

where R_{11} and R_{13} are as in (6). By changing the variable of integration in the second part of the integral by setting $y' = -z'$ and then writing z' for y', the following alternative form is obtained:

$$\int_{-h+\delta}^{h-\delta} I_{1z}^{(a)}(z') K_A^{(a)}(z,z') \, dz'$$

$$= -\frac{j4\pi}{\zeta_o} \left[C_1 \cos \beta_o z + C_2 \sin \beta_o z - \theta_A^{(a)}(z) \right], \qquad (90b)$$

where

$$K_A^{(a)}(z, z') = \frac{e^{-j\beta_0 R_{11}}}{R_{11}} + \frac{e^{-j\beta_0 R_{13}'}}{R_{13}'} \qquad (91a)$$

$$R_{11} = \sqrt{(z - z')^2 + a^2}; \quad R_{13}' = \sqrt{(z + z')^2 + (2h)^2}.$$
(91b)

Similarly, the scalar potential $\phi_{1x}(z)$ maintained on conductor 1 by the charges on conductors 2 and 4 is obtained from (10) with (89c). By changing the variable of integration to $y' = -x'$ and then writing x' for y', the following integral is obtained:

$$\phi_{1x}^{(a)}(z) = \frac{1}{4\pi\epsilon_o} \int_{-h+\delta}^{h-\delta} q_2^{(a)}(x') K_B^{(a)}(z, x') \, dx',$$
(92)

where

$$K_B^{(a)}(z, x') = \frac{e^{-j\beta_0 R_{12}}}{R_{12}} - \frac{e^{-j\beta_0 R_{14}'}}{R_{14}'}$$
(93a)

and

$$R_{12} = \sqrt{(h + z)^2 + (h + x')^2 + a^2},$$
$$R_{14}' = \sqrt{(h - z)^2 + (h - x')^2 + a^2}. \quad (93b)$$

It is now convenient to change the variables so that the origin of the new temporary coordinates u and v is at the corner of the square with coordinates $(-h, -h)$. That is, let

$$v = h + z; \quad v' = h + z'; \quad u = h + x;$$
$$u' = h + x'. \quad (94)$$

In terms of these coordinates (90) through (93) become

$$\int_{\delta}^{2h-\delta} I_{1z}^{(a)}(v') K_A^{(a)}(v, v') \, dv'$$
$$= \frac{-j4\pi}{\zeta_o} [C_3 \cos \beta_o v + C_4 \sin \beta_o v - \theta_A^{(a)}(v)], \quad (95)$$

$$K_A^{(a)}(v, v') = \frac{e^{-j\beta_0 R_{11}}}{R_{11}} + \frac{e^{-j\beta_0 R_{13}'}}{R_{13}'},$$
(96a)

$$R_{11} = \sqrt{(v - v')^2 + a^2},$$
$$R_{13}' = \sqrt{(v + v' - 2h)^2 + (2h)^2}; \quad (96b)$$

$$\phi_{1x}^{(a)}(v) = \frac{1}{4\pi\epsilon_o} \int_{\delta}^{2h-\delta} q_2(u') K_B^{(a)}(v, u') \, du'$$
(97)

$$K_B^{(a)}(v_1, u_c') = \frac{e^{-j\beta_0 R_{12}}}{R_{12}} - \frac{e^{-j\beta_0 R_{14}'}}{R_{14}'},$$
(98a)

$$R_{12} = \sqrt{v^2 + u'^2 + a^2},$$
$$R_{14}' = \sqrt{(2h - v)^2 + (2h - u')^2 + a^2}. \quad (98b)$$

The new arbitrary constants C_3 and C_4 are related to C_1 and C_2 by

$$C_3 = C_1 \cos \beta_o h - C_2 \sin \beta_o h;$$
$$C_4 = C_1 \sin \beta_o h + C_2 \cos \beta_o h.$$

The expression for the scalar potential on conductor 1 is

$$\phi_1^{(a)}(v) = -C_3 \sin \beta_o v + C_4 \cos \beta_o v + \theta_v^{(a)}(v).$$
(99a)

The driving voltage V_{12} as given by (23) is

$$V_{12}^{(a)} = \phi_1^{(a)}(v = \delta) - \phi_2^{(a)}(u = \delta) = 2\phi_1^{(a)}(v = \delta).$$
(99b)

The substitution of (99b) in (99a) with $v = \delta$ gives

$$C_4 = C_4' - \frac{\theta_V^{(a)}(v = \delta)}{\cos \beta_o \delta};$$

$$C_4' = \frac{1}{2} \frac{V_{12}^{(a)} - \theta_V^{(a)}(v = \delta)}{\cos \beta_o \delta}, \quad (100)$$

where C_4' is introduced for later convenience.

The functions $\theta_A^{(a)}(v)$ and $\theta_V^{(a)}(v = \delta)$ in (95) and (100) are obtained from (15) with appropriate changes in variable. Thus, with $z_o = \delta$,

$$\theta_A^{(a)}(v) = \beta_o \int_{\delta}^{v} \phi_{1x}^{(a)}(w) \cos \beta_o(v - w) \, dw$$
$$- \phi_{1x}(\delta) \sin \beta_o(v - \delta); \quad (101a)$$

$$\theta_V^{(a)}(v) = \int_{\delta}^{v} \phi_{1x}(w) \sin \beta_o(v - w) \, dw$$
$$+ \phi_{1x}(\delta) \cos \beta_o(v - \delta). \quad (101b)$$

With $v = \delta$, this last expression reduces to

$$\theta_V^{(a)}(v = \delta) = \phi_{1x}^{(a)}(\delta).$$
(101c)

The substitution of (100) with (101c) in (95) gives the following integral equation:

$$\int_{\delta}^{2h-\delta} I_{1z}^{(a)}(v') K_A^{(a)}(v, v') \, dv'$$
$$= \frac{-j4\pi}{\zeta_o} \left\{ [C_3 \cos \beta_o v + C_4' \sin \beta_o v] \right.$$
$$\left. \left[\theta_A^{(a)}(v) + \phi_{1x}^{(a)}(\delta) \frac{\sin \beta_o v}{\cos \beta_o \delta} \right] \right\}. \quad (102)$$

This equation resembles that for the isolated straight antenna[11], [12] and may be solved in the same manner. Thus, let

$$g^{(a)}(v, v') = I_{1z}^{(a)}(v')/I_{1z}^{(a)}(v), \quad (103a)$$

and

$$\Psi^{(a)}(v) = \int_{\delta}^{2h-\delta} g^{(a)}(v, v') K_A^{(a)}(v, v') \, dv'$$
$$= \Psi^{(a)} + \gamma^{(a)}(v), \quad (103b)$$

where $\Psi^{(a)}$ is the sensibly constant magnitude of $\Psi^{(a)}(v)$. As in the case of the V-antenna and the center-driven antenna,

$$\Psi^{(a)} = \begin{cases} \Psi^{(a)}(h - \delta - \lambda/4|, & \beta_o(h - \delta) \geqq \pi/2 \\ \Psi^{(a)}(0)|/\sin \beta_o(h - \delta), & \beta_o(h - \delta) = \pi/2 \end{cases}.$$
(103c)

The transposition of the left-hand member in (102) to the right side and the simultaneous addition of
$$I_{1z}^{(a)}(v)\Psi^{(a)}(v) = I_{1z}^{(a)}(v)[\Psi^{(a)} + \gamma^{(a)}(v)]$$
leads to the following equivalent equation:

$$I_{1z}^{(a)}(v) = \frac{-j4\pi}{\zeta_o \Psi^{(a)}} \left\{ [C_3 \cos \beta_o v + C_4' \sin \beta_o v] \right.$$
$$\left. -\left[\theta_A^{(a)}(v) + \phi_{1x}^{(a)}(\delta) \frac{\sin \beta_o v}{\cos \beta_o \delta} \right] \right\}$$
$$- \frac{1}{\Psi^{(a)}} [D(v) + I_{1z}^{(a)}(v)\gamma^{(a)}(v)], \quad (104a)$$

where

$$D(v) = \int_{\delta}^{2h-\delta} [I_{1z}{}^{(a)}(v')$$
$$- I_{1z}{}^{(a)}(v)g^{(a)}(v, v')]K_A{}^{(a)}(v, v')\, dv'. \quad (104b)$$

Since in the diagonal dipole mode $I_{1z}{}^{(a)}(v = 2h - \delta) = 0$, this quantity may be subtracted from both sides of (104a).

$$I_{1z}{}^{(a)}(v) = \frac{-j4\pi}{\zeta_o \Psi^{(a)}} \Big\{ [C_3 F_{ov}{}^{(a)} + C_4' G_{ov}{}^{(a)}]$$
$$- \Big[\theta_{Av}{}^{(a)} + \phi_{1x}{}^{(a)}(\delta) \frac{G_{ov}{}^{(a)}}{\cos \beta_o \delta} \Big] \Big\}$$
$$- \frac{1}{\Psi^{(a)}} [D_v + I_{1z}{}^{(a)}(v)\gamma^{(a)}(v)], \quad (105)$$

where

$$F_{ov}{}^{(a)} = F_o(v) - F_o(2h - \delta)$$
$$= \cos \beta_o v - \cos \beta_o(2h - \delta), \quad (106a)$$
$$G_{ov}{}^{(a)} = G_o(v) - G_o(2h - \delta)$$
$$= \sin \beta_o v - \sin \beta_o(2h - \delta), \quad (106b)$$
$$\theta_{Av}{}^{(a)} = \theta_A{}^{(a)}(v) - \theta_A{}^{(a)}(v = 2h - \delta), \quad (106c)$$
$$D_v = D(v) - D(v = 2h - \delta)$$
$$= D(v) - \int_{\delta}^{2h-\delta} I_{1z}{}^{(a)}(v')K_A{}^{(a)}(2h - \delta, v')\, dv'. \quad (106d)$$

The zeroth-order current in (105) is

$$[I_{1z}{}^{(a)}(v)]_o = \frac{-j4\pi}{\zeta_o \Psi^{(a)}} [C_3 F_{ov}{}^{(a)} + C_4' G_{ov}{}^{(a)}]. \quad (107$$

The zeroth-order charge per unit length is obtained using the equation of continuity,

$$\frac{dI(v)}{dv} + j\omega q(v) = 0.$$

It is

$$[q_1{}^{(a)}(v)]_o = \frac{4\pi\epsilon_o}{\Psi^{(a)}} [-C_3 G_o(v) + C_4' F_o(v)]. \quad (108a)$$

With (89c) it follows that

$$[q_1{}^{(a)}(u)]_o = \frac{4\pi\epsilon_o}{\Psi^{(a)}} [C_3 G_o(u) - C_4' F_o(u)]. \quad (108b)$$

If this expression is substituted in (97), the following first-order formula is obtained for $\phi_{1x}{}^{(a)}(v)$:

$$[\phi_{1x}{}^{(a)}(v)]_1 = - \frac{1}{\Psi^{(a)}} [-C_3 J_{1s}{}^{(a)}(v) + C_4' J_{1c}{}^{(a)}(v)], \quad (109)$$

where

$$J_{1s}{}^{(a)}(v) = - \int_{\delta}^{2h-\delta} \sin \beta_o u' K_\beta{}^{(a)}(v, u')\, du', \quad (110)$$
$$J_{1c}{}^{(a)}(v) = \int_{\delta}^{2h-\delta} \cos \beta_o u' K_\beta{}^{(a)}(v, u')\, du'. \quad (111)$$

The first-order formula for $\theta_A{}^{(a)}(v)$ is obtained by using (109) in (101a). Thus,

$$[\theta_A{}^{(a)}(v)]_1 = \beta_o \int_{\delta}^{v} [\phi_{1x}{}^{(a)}(w)]_1 \cos \beta_o(v - w)\, dw$$
$$- \phi_{1x}{}^{(a)}(\delta) \sin \beta_o(v - \delta). \quad (112)$$

With (109) this becomes

$$[\theta_A{}^{(a)}(v)]_1 = - \frac{1}{\Psi^{(a)}} [C_3 L_{1s}(v) + C_4' L_{1c}{}^{(a)}(v)], \quad (11?)$$

where

$$L_{1s}{}^{(a)}(v) = \int_{\delta}^{v} J_{1s}{}^{(a)}(w) \cos \beta_o(v - w)\beta_o\, dw$$
$$- J_{1s}{}^{(a)}(\delta) \sin \beta_o(v - \delta) \quad (11\cdot$$
$$L_{1c}{}^{(a)}(v) = \int_{\delta}^{v} J_{1c}{}^{(a)}(w) \cos \beta_o(v - w)\beta_o\, dw$$
$$- J_{1c}{}^{(a)}(\delta) \sin \beta_o(v - \delta). \quad (11!$$

If the zeroth-order current (107) is substituted in th integrals in (105) a first-order solution is obtained in th form,

$$[I_{1z}{}^{(a)}(v)]_1$$
$$= \frac{-j4\pi}{\zeta_o \Psi^{(a)}} \Big\{ C_3 \Big[F_{ov}{}^{(a)} + \frac{S_{1v}}{\Psi^{(a)}} \Big\} + C_4' \Big[G_{ov}{}^{(a)} + \frac{T_{1v}}{\Psi^{(a)}} \Big] \Big\}, (11\,$$

where

$$S_{1v} = S_1(v) - S_1(2h - \delta);$$
$$T_{1v} = T_1(v) - T_1(2h - \delta); \quad (11$$

and

$$S_1(v) = F_1{}^{(a)}(v) + L_{1c}{}^{(a)}(v)$$
$$+ J_{1c}{}^{(a)}(\delta) F_0(v)/\cos \beta_o \delta; \quad (118$$
$$T_1(v) = G_1{}^{(a)}(v) + L_{1s}{}^{(a)}(v)$$
$$+ J_{1s}{}^{(a)}(\delta) G_0(v)/\cos \beta_o \delta. \quad (118$$

The functions $S_1(2h - \delta)$ and $T_1(2h - \delta)$ are obtain from (118a, b) by setting $v = 2h - \delta$. Note that $L_{1s}{}^{(a)}($ and $L_{1c}{}^{(a)}(v)$ are defined in (114) and (115). Also,

$$F_1{}^{(a)}(v) = - \int_{\delta}^{2h-\delta} [F_{0v'}{}^{(a)}$$
$$- F_{ov}{}^{(a)}g^{(a)}(v, v')K_A{}^{(a)}(v, v')\, dv' - F_{0v}{}^{(a)}\gamma^{(a)}(v), \quad (119$$
$$F_1{}^{(a)}(2h - \delta)$$
$$= - \int_{\delta}^{2h-\delta} F_{ov'}{}^{(a)}K_A{}^{(a)}(2h - \delta, v')\, dv'. \quad (119\cdot$$

$G_1{}^{(a)}(v)$ and $G_1{}^{(a)}(2h - \delta)$ are given by (119a) and (119 with G substituted for F.

With (100), the first-order current (116) becomes

$$[I_{1z}{}^{(a)}(v)]_1 = \frac{-j4\pi}{\zeta_o \Psi^{(a)} \cos \beta_o \delta} \big\{ C_3 [(F_{ov}{}^{(a)} + S_{1v}/\Psi^{(a)})$$
$$\times \cos \beta_o \delta + (G_{ov}{}^{(a)} + T_{1v}/\Psi^{(a)}) \sin \beta_o \delta]$$
$$+ \tfrac{1}{2} V_{12}{}^{(a)}(G_{ov}{}^{(a)} + T_{1v}/\Psi^{(a)}) \big\}. \quad (12$$

The arbitrary constant C_3 may be evaluated fro (104a) by setting $v = 2h - \delta$ and $I_{1z}(v = 2h - \delta) =$ The resulting equation is

$$C_3 F_o{}^{(a)}(2h - \delta) + C_4' G_o{}^{(a)}(2h - \delta) - \theta_A{}^{(a)}(2h - \delta)$$
$$- \frac{\phi_{1x}(\delta)G_o{}^{(a)}(2h - \delta)}{\cos \beta_o \delta} + D(2h - \delta) = 0. \quad (12$$

The substitution of first-order values for $\theta_A{}^{(a)}(2h -$ $\theta_{1x}(\delta)$ and $D(2h - \delta)$ gives

$$C_3\left[F_o^{(a)}(2h - \delta) + \frac{S_1(2h - \delta)}{\Psi^{(a)}}\right]$$

$$+ \left[\frac{\frac{1}{2}V_{12}^{(a)} + C_3 \sin \beta_o \delta}{\cos \beta_o \delta}\right]\left[G_o^{(a)}(2h - \delta)\right.$$

$$\left. + \frac{T_1(2h - \delta)}{\Psi^{(a)}}\right] = 0, \quad (122)$$

where $S_1(2h - \delta)$ and $T_1(2h - \delta)$ are given by (118a, b) with $v = 2h - \delta$.

The solution of (122) for C_3 results in

$$C_3 = -\frac{V_{12}^{(a)}}{2D}\left[G_o^{(a)}(2h - \delta) + \frac{T_1(2h - \delta)}{\Psi^{(a)}}\right]$$

where

$$D = \left[F_o^{(a)}(2h - \delta) + \frac{S_1(2h - \delta)}{\Psi^{(a)}}\right]\cos \beta_o \delta$$

$$+ \left[G_o^{(a)}(2h + \delta) + \frac{T_1(2h - \delta)}{\Psi^{(a)}}\right]\sin \beta_o \delta. \quad (123)$$

If (123) is substituted in (120) and the terms are re-arranged, and simplified the following result is obtained. The temporary variable v is removed by the substitution, $v = z + h$

$$[I_{1z}^{(a)}(z)]_1 = \frac{j2\pi V_{12}^{(a)}}{\zeta_o \Psi^{(a)}}$$

$$\times \left\{\frac{\sin \beta_o(h - \delta - z) + M_1^{(a)}(h + z)/\Psi^{(a)}}{\cos 2\beta_o(h - \delta) + A_1^{(a)}/\Psi^{(a)}}\right\}, \quad (124)$$

where

$$M_1^{(a)}(h + z) = S_1(h + z) \sin \beta_o(2h - \delta)$$

$$- T_1(h + z) \cos \beta_o(2h - \delta) + T_1(2h - \delta) \cos \beta_o(h + z)$$

$$- S_1(2h - \delta) \sin \beta_o(h + z). \quad (125a)$$

$$A_1^{(a)} = S_1(2h - \delta) \cos \beta_o \delta + T_1(2h - \delta) \sin \beta_o \delta \quad (125b)$$

and where $(h - \delta) \leqq z \leqq (-h + \delta)$.

The corresponding admittance is obtained by setting $z = -h + \delta$. It is

$$[Y^{(a)}]_1 = \frac{j2\pi}{\zeta_o \Psi^{(a)}}\left[\frac{\sin 2\beta_o(h - \delta) + M_1^{(a)}(\delta)/\Psi^{(a)}}{\cos 2\beta_o(h - \delta) + A_1^{(a)}/\Psi^{(a)}}\right] \quad (126)$$

Note that the zeroth-order impedance is

$$[Z^{(a)}]_o = \frac{-j\zeta_o \Psi^a}{2\pi} \cot 2\beta_o(h - \delta). \quad (127)$$

The currents in the first- and third-phase sequences are obtained by substituting (124) in (88e) and (88f) and replacing $V_{12}^{(a)}$ by $V^{(1)}$ and $V^{(3)}$ respectively. Thus

$$[I_{1z}^{(1)}(z)]_1 = [I_{1z}^{(a)}(z) - j[I_{1z}^{(a)}(-z)] = \frac{j2\pi V^{(1)}}{\zeta_o \Psi^{(a)}}$$

$$\times \left\{\frac{\sin \beta_o(h - \delta - z) - j \sin \beta_o(h - \delta + z)}{\cos 2\beta_o(h - \delta) + A_1^{(a)}/\Psi^{(a)}}\right\}; \quad (128)$$

$$[I_{1z}^{(3)}(z)]_1 = [I_{1z}^{(a)}(z) + j[I_{1z}^{(a)}(-z)] = \frac{j2\pi V^{(3)}}{\zeta_o \Psi^{(a)}}$$

$$\times \left\{\frac{\sin \beta_o(h - \delta - z) + j \sin \beta_o(h - \delta + z)}{\cos 2\beta_o(h - \delta) + A_1^{(a)}/\Psi^{(a)}}\right\}. \quad (129)$$

Since $I_{1z}^{(a)}(h - \delta) = 0$, the corresponding admittances are simply

$$[Y^{(1)}]_1 = [Y^{(3)}]_1 = [Y^{(a)}]_1, \quad (130)$$

where $[Y_1^{(a)}]_1$, is in (126). The corresponding zeroth-order impedances $[Z^{(1)}]_o$ and $[Z^{(3)}]_o$ are like (127).

The function $\Psi^{(a)}$ may be evaluated by substituting the zeroth-order current from (124) in (103b). Thus, $g^{(a)}(v, v') = \sin \beta_o(2h - \delta - v')/\sin \beta_o(2h - \delta - v)$. If the small quantity δ is neglected in the evaluation of $\Psi^{(a)}$, it follows that

$$\Psi^{(a)}(v) \doteq \Psi^{(a)} + \gamma^{(c)}(v)$$

$$\doteq \int_o^{2h} \sin \beta_o(2h - v') K_A^{(a)}(v, v') \, dv'$$

$$= \sin 2\beta_o h \int_o^{2h} \cos \beta_o v' K_A^{(a)}(v, v') \, dv'$$

$$- \cos 2\beta_o h \int_o^{2h} \sin \beta_o v' K_A^{(a)}(v, v') \, dv'. \quad (131a)$$

These integrals can be expressed in terms of generalized sine- and cosine-integral functions and the magnitude $\Psi^{(a)}$ determined. As in the case of the V-antenna[7] and the center-driven straight antenna[11, 12]

$$\Psi^{(a)} = \begin{cases} |\Psi^{(a)}(2h - \lambda_o/4)| & \text{when } 2\beta_o h \geqq \pi/2 \\ |\Psi^{(a)}(0)| /\sin 2\beta_o h & \text{when } 2\beta_o h \leqq \pi/2. \end{cases} \quad (131b)$$

The functions S_1 and T_1 which occur in (126) and (130) involve functions of the same type as encountered in the first-order terms of the zeroth- and second-phase sequences. They may be evaluated numerically.

VI. SUPERPOSITION OF ALL FOUR PHASE SEQUENCES—THE GENERAL SOLUTION

Since the driving voltages at the four corners may all be different, or up to three of them may be replaced by tuning or load impedances, there are many combinations involving the superposition of all four phase sequences. In general these are best determined by first evaluating the individual phase-sequence currents and then combining these. Only two relatively simple and important cases are considered here.

Loop Driven at One Corner

When all four phase-sequence voltages are equal so that

$$V^{(0)} = V^{(1)} = V^{(2)} = V^{(3)}, \quad (132a)$$

it follows that

$$V_{12} = 4V^{(0)}; \quad V_{23} = V_{34} = V_{41} = 0. \quad (132b)$$

This corresponds to the loop driven only at one corner with each of the other three corners short circuited. The currents in conductors 1 and 4 are given by (26a) and (26d). With (128) and (129) they are

$$I_{1z}(z) = I_{1z}{}^{(0)}(z) + I_{1z}{}^{(2)}(z) + I_{1z}{}^{(1)}(z) + I_{1z}{}^{(3)}(z)$$
$$= I_{1z}{}^{(0)}(z) + I_{1z}{}^{(2)}(z) + 2I_{1z}{}^{(a)}(z), \quad (133a)$$

$$I_{4x}(x) = I_{4x}{}^{(0)}(x) - I_{4x}{}^{(2)}(x) - jI_{4x}{}^{(1)}(x) + jI_{4x}{}^{(3)}(x)$$
$$= I_{4x}{}^{(0)}(x) - I_{4x}{}^{(2)}(x) - 2I_{4x}{}^{(a)}(-x), \quad (133b)$$

where the four phase sequence currents in the first-order are in (49), (82), (128) and (129). $I_{1z}{}^{(a)}(z)$ is in (124); in this case $V_{12}{}^{(a)} = V^{(1)} = V^{(3)}$. There is no gain in simplicity by combining the four formulas into a single one. But the zeroth-order currents may be combined and simplified. Thus,

$$[I_{1z}(z)]_o = \frac{j2\pi V^{(0)}}{\zeta_o} \left\{ \frac{\cos \beta_o z}{\Psi^{(0)} \sin \beta_o(h - \delta)} \right.$$
$$\left. - \frac{\sin \beta_o z}{\Psi^{(2)} \cos \beta_o(h - \delta)} + \frac{\sin \beta_o(h - \delta - z)}{\Psi^{(a)} \cos 2\beta_o(h - \delta)} \right\}, \quad (134a)$$

$$[I_{4x}(x)]_o = \frac{j2\pi V^{(0)}}{\zeta_o} \left\{ \frac{\cos \beta_o x}{\Psi^{(0)} \sin \beta_o(h - \delta)} \right.$$
$$\left. + \frac{\sin \beta_o x}{\Psi^{(2)} \cos \beta_o(h - \delta)} - \frac{2 \sin \beta_o(h - \delta - x)}{\Psi^{(a)} \cos 2\beta_o(h - \delta)} \right\}. \quad (134b)$$

Although the three Ψ-functions that occur in these expressions are not the same, they do not differ greatly and a qualitative picture of the distributions of current may be obtained by replacing them by a mean value Ψ. The leading terms in the zeroth-order currents are then found to be:

$$[I_{1z}(z)]_o = -\frac{j2\pi V^{(0)}}{\zeta_o \Psi} \left[\frac{\cos \beta_o(3h - 3\delta - z)}{\sin 4\beta_o(h - \delta)} \right];$$
$$(-h + \delta \leqq z \leqq h - \delta); \quad (135a)$$

$$[I_{4x}(x)]_o = \frac{-j2\pi V^{(0)}}{\zeta_o \Psi} \left[\frac{\cos \beta_o(h - \delta - x)}{\sin 4\beta_o(h - \delta)} \right];$$
$$(-h + \delta \leqq x \leqq h - \delta). \quad (135b)$$

The zeroth-order driving-point impedance is

$$[Z_{in}]_o = \frac{V^{(0)}}{[I_{1z}(h - \delta)]_o} = \frac{j\zeta_o \Psi}{2\pi} \tan 4\beta_o(h - \delta). \quad (136)$$

This is the input impedance of a lossless short-circuited transmission line of electrical length $4\beta_o(h - \delta)$ and characteristic impedance $\zeta_o\Psi/2\pi$.

Square Rhombic Antenna

The square rhombic antenna is driven at one corner and loaded at the diagonally opposite corner with a resistance so chosen that the distribution of current around the loop approximates a travelling wave. In order to investigate this problem let

$$V_{12} \neq 0; \quad V_{23} = V_{41} = 0;$$
$$V_{34} = -I_{4x}(h - \delta)Z_{34} ; \quad (137)$$

here Z_{34} is an arbitrary impedance to be chosen so that the current distribution approximates a traveling wave. The phase-sequence voltages obtained from (25a–d) with (137) are:

$$V^{(2)} = V^{(0)} = \tfrac{1}{4}[V_{12} + V_{34}]$$
$$= \tfrac{1}{4}[V_{12} - I_{4x}(h - \delta)Z_{34}], \quad (138a)$$

$$V^{(3)} = V^{(1)} = \tfrac{1}{4}[V_{12} - V_{34}]$$
$$= \tfrac{1}{4}[V_{12} + I_{4x}(h - \delta)Z_{34}]. \quad (138b)$$

The use of (138a) in (49) and (82), and of (138b) i (128) and (129) provides the four phase-sequence cu rents for substitution in (26a–d) or (133a) and (133b Since the first-order phase-sequence currents yield complicated sum that is not readily simplified in genera it is instructive to examine the zeroth-order current The leading term in the zeroth-order sum is obtaine using a mean value Ψ for the three slightly different functions. With this simplification the approxima zeroth-order currents in conductors 1 and 4 are

$$[I_{1z}(z)]_o \doteq \frac{-j2\pi}{\zeta_o\Psi} \times$$
$$\frac{V_{12} \cos \beta_o(3h - 3\delta - z) + V_{34} \cos \beta_o(h - \delta + z)}{\sin 4\beta_o(h - \delta)}, \quad (139a)$$

$$[I_{4x}(x)]_o \doteq \frac{-j2\pi}{\zeta_o\Psi} \times$$
$$\frac{V_{12} \cos \beta_o(h - \delta - x) + V_{34} \cos \beta_o(3h - 3\delta + x)}{\sin 4\beta_o(h - \delta)}, \quad (139b)$$

where $V_{34} = -[I_{4x}(h - \delta)]_o[Z_{34}]_o$. If this is substitute in (139b) for $x = h - \delta$ and the resulting expression solved for $[I_{4x}(h - \delta)]_o$, the result is

$$[I_{4x}(h - \delta)]_o$$
$$= V_{12} \left\{ [Z_{34}]_o \cos 4\beta_o(h - \delta) + j\frac{\zeta_o\Psi}{2\pi} \sin 4\beta_o(h - \delta) \right\}^{-1}$$
$$(140)$$

It is readily verified that if

$$[Z_{34}]_o = \frac{\zeta_o\Psi}{2\pi} = 60\Psi \text{ ohms} \quad (141)$$

that (140) reduces to

$$[I_{4x}(h - \delta)]_o = \frac{2\pi}{\zeta_o\Psi} V_{12}e^{-j4\beta_o(h-\delta)} \quad (142)$$

which is the appropriate form for the current at $x =$ $h - \delta$ if the distribution along the conductors is traveling wave.

With (141) and (142) it follows that $V_{34} = V_{12}e^{-j4\beta_o(h-\delta)}$. If this value is inserted in (139a) and (139b) the result are,

$$[I_{1z}(z)]_o = \frac{2\pi V_{12}}{\zeta_o\Psi} e^{-j\beta_o(h-\delta+z)} ;$$
$$(-h + \delta \leqq z \leqq h - \delta); \quad (143a)$$

$$[I_{4x}(x)]_o = \frac{2\pi V_{12}}{\zeta_o\Psi} e^{-j\beta_o(3h-3\delta+x)} ;$$
$$(-h + \delta \leqq x \leqq h - \delta). \quad (143b)$$

The approximate input impedance is

$$Z_{in} = \frac{\zeta_o\Psi}{2\pi} = 60\Psi \text{ ohms.} \quad (144)$$

Eqs. (143a, b) indicate a traveling wave along conductor 1 from the generator at $z = -h + \delta$ to its end at $z = h - \delta$ where it is connected by a short-circuit (e.g.,

quarter-wavelength of open-end two-wire line) to the beginning of conductor 4 and $x = -h + \delta$. The traveling wave continues along conductor 4 to its end at $x = h - \delta$ where it is connected to one side of a two-wire line of length $\lambda/2$ terminated in $Z_{34} = \zeta_o \Psi/2\pi$. A similar traveling wave exists simultaneously on conductors 2 and 3 but with the current instantaneously in opposite directions at corresponding points.

A more accurate determination of the distribution of current and the input impedance of a square rhombic antenna may be achieved by evaluating the first-order formulas corresponding to the approximate zeroth-order results (143a, b).

Conclusion

First-order formulas for the distribution of current and the driving point impedance of a square loop antenna have been derived for the general case involving a generator or an arbitrary impedance at each corner. Account has been taken of the finite cross section of the conductors and no restriction has been imposed on the dimension of the square although it is to be expected that as in a similar analysis of the center-driven cylindrical antenna[11, 12] the results are quantitatively accurate only for squares with sides that do not exceed two-wavelengths. Although conductors of circular cross section have been assumed, the results may be extended to other shapes by determining the radius of the equivalent circle.[12] The analysis has been carried out assuming all generators and loads to be connected to the corners in a manner equivalent to Fig. 6. Attention has been called to the need of lumped junction networks at the corners to take account of transmission-line end effects and coupling of the line and the loop. Accurate results cannot be expected unless such junction networks are included. It was stated at the outset that circuits in which currents on one or more transmission lines contribute to radiation are excluded. These do not properly constitute loop antennas but combinations of a loop with linear elements. Such circuits occur with the diagonal dipole mode when conductors of any kind are connected to extend the physical size of the charged pair of corners. Specifically in Fig. 15(a), no external sections of transmission line may be attached to the upper left- or lower right-hand corner if the two generators are active.

It is anticipated that zeroth-order distributions of current are satisfactory for calculating radiation fields, but that at least first-order input currents are required for determining quantitatively adequate driving-point impedances.

The Radiation Pattern and Induced Current in a Circular Antenna with an Annular Slit

JOSEF MEIXNER†

Summary—A finite plane antenna is considered which has holes on one side that act as sources of radiation and which is on the remaining parts of this side and on the whole other side perfectly conducting. The purpose of this paper is to develop an approximation method for the computation of the radiation pattern which works well if the finite plane and the distance of the holes from its boundary are large compared with the wave length. This is achieved by computing the radiation field of a corresponding infinite plane antenna and subtracting from it the field produced by the current induced in the infinite plane outside the finite plane antenna. Numerical results for the circular antenna with annular slit show that this approximation method is very satisfactory.

INTRODUCTION

AN INFINITE perfectly conducting plane has on one side a generator of electromagnetic waves which propagate through one or more holes in the plane to the other half-space and give there a radiation pattern. We call this setup a plane antenna, the holes acting as the sources of the radiation. In many cases the electromagnetic field in the holes is approximately known and then the electromagnetic field in the second half-space can be expressed in integral form from its values in the holes. Actually it suffices to know only the components of the electric field strength parallel to the plane. Important cases are a coaxial line ending in the plane or a slit between whose rims an alternating voltage is applied.

When the infinite plane is reduced to a finite, but very large, screen, one expects that the radiation pattern of the antenna is not much changed. For simplicity we assume that the generator in the rear of the plane is screened in such a way that it does not radiate directly into its own half-space. A backward radiation arises therefore only through the diffraction of the waves around the boundary of the screen. In mathematical language this problem can be idealized and formulated as a boundary value problem for a plane infinitely thin screen with the tangential component of the electric field strength vanishing everywhere on one side, while it assumes prescribed values in the "holes" on the other side and vanishes on the rest of this side of the screen. The apparatus driving this antenna is assumed to be located between the two sides of the screen and therefore to be "infinitely thin." This is a not too severe restriction if the screen is large compared with the wavelength and the holes are sufficiently far from its boundary.

It is hardly possible to compute rigorously the radiation patterns of such a finite plane antenna of arbitrary shape. If it is large enough the radiation pattern on one side will be given at least approximately by replacing the screen by an infinite plane. But even then the radiation in directions near the plane will be badly approximated while the backward radiation is entirely neglected.

A tentative way to find a better approximation for the radiation pattern of the finite screen starts from an approximation of the current distribution. For the infinite plane it can be easily expressed in integral form. For the finite screen we take the same current distribution cutoff at the rim of the screen and thus neglect the disturbance of the current distribution caused by this cutting off process. It is well known that the radiation pattern and the power radiated by the antenna are not very sensitive to errors in the assumed form of the current distribution and so we can expect that such a cut off current distribution gives good results even for the backward scattering.[1]

The purpose of this paper is to show for a special case, the circular antenna with a circular slit, how this idea works and so to give an approximation method which will perhaps prove useful in other cases.

A GENERAL THEOREM

As outlined in the Introduction, we consider three different plane antennas located in the x, y-plane. The first is the infinite plane radiating into the half-space $z \geq 0$, the second is the double-sided screen S with its "holes" on the side $z = +0$, the third is the screen with the current of the infinite plane antenna cut off.

From the values of the x- and, y-component of the electric field strength in the whole plane $z = 0$ one can compute the magnetic field H by an integral representation which we denote by

$$\vec{H} = \pm [I(S) + I(C)]\vec{E}$$

$$(z \geq 0 \text{ and } z \leq 0 \text{ respectively}). \qquad (1)$$

The operator $I(S)$ contains an integral over the screen S, while in $I(C)$ the integration is to be performed over the part of the plane exterior to the screen, or the complementary screen C. Further, we distinguish between S_+ and S_-, C_+ and C_-, according to whether the

---

† Dept. of Mathematics, Michigan State Univ.; on leave of absence from Technische Hochschule, Aachen, Germany.

[1] S. A. Schelkunoff, "Antennas: Theory and Practice," p. 214; 1952.

integration is to be carried out over the positive or the negative x, y-plane.

For the first antenna we get, because E_x and E_y vanish on C and S_-,

$$\overrightarrow{H_1} = I(S_+)\overrightarrow{E} \quad (z \geqq 0)$$
$$= 0 \quad (z \leqq 0) . \tag{2}$$

For the second antenna we compute the magnetic field $\overrightarrow{H_2}$ in the x, y-plane in two ways, from the values of E_x, E_y on C_+ and S_+ and on C_- and S_-. As the values on C_+ and C_- do not differ and as E_x and E_y vanish on S_-, we get

$$\overrightarrow{H_2} = [I(S_+) + I(C)]\overrightarrow{E} = -I(C)\overrightarrow{E},$$
$$(z = \pm0; x, y \text{ in } C). \tag{3}$$

For the third antenna, we have in addition to $\overrightarrow{H_1}$ a contribution which comes from the cutoff current: it is $\overrightarrow{H'} + \overrightarrow{H''}$ on C_+ and $-\overrightarrow{H'} + \overrightarrow{H''}$ on C_- with $\overrightarrow{H'}$ parallel to the x, y-plane and $\overrightarrow{H''}$ parallel to the z-axis. The total field is therefore $I(S_+)E + H' + H''$ on C_+ and $-H' + H''$ on C_-. It must be continuous when passing through C, the complementary screen C being removed now, and therefore we have on C

$$I(S_+)\overrightarrow{E} = -2\overrightarrow{H'}. \tag{4}$$

Calling the resultant field in C for the latter case $\overrightarrow{H_3}$, we find from (2), (3), and (4) the following relation between the magnetic fields of the three antennas:

$$\overrightarrow{H_2} = \overrightarrow{H_3} - \overrightarrow{H''} = \tfrac{1}{2}\overrightarrow{H_1} \quad (z = +0, x, y \text{ in } C). \tag{5}$$

A similar statement holds for the radiation patterns in the direction of the plane $z = 0$.

We remark that generally $H_{1z} = H_{2z} = 0$ on C. If S is circular in shape and the current distribution has only a radial component, then also H'' vanishes on C.

For all the antennas we have assumed the same electric field in the holes.

The Circular Antenna with an Annular Slit

We consider now an antenna of the first type with a very narrow slit. x, y are Cartesian, ϱ, φ polar co-ordinates in the plane of the slit, r, ϑ, φ are spherical co-ordinates with $\vartheta = 0$ as positive z-axis, and $\varrho = \varrho_o$ is the slit. Between its rims an alternating tension U with frequency ω (wave number k) is applied. The phase and amplitude of U do not depend on φ.

With the help of the Rayleigh integral one gets, after some manipulations,[2,3] the following expression for the distant magnetic field:

[2] W. Kloepfer, "Theorie der Kreisscheibenantennen," Thesis, Aachen; 1950.

[3] J. Meixner and W. Kloepfer, "Theorie der ebenen ringspalt-antenne," Z. angew. Phys., vol. 3, pp. 171–178; 1951.
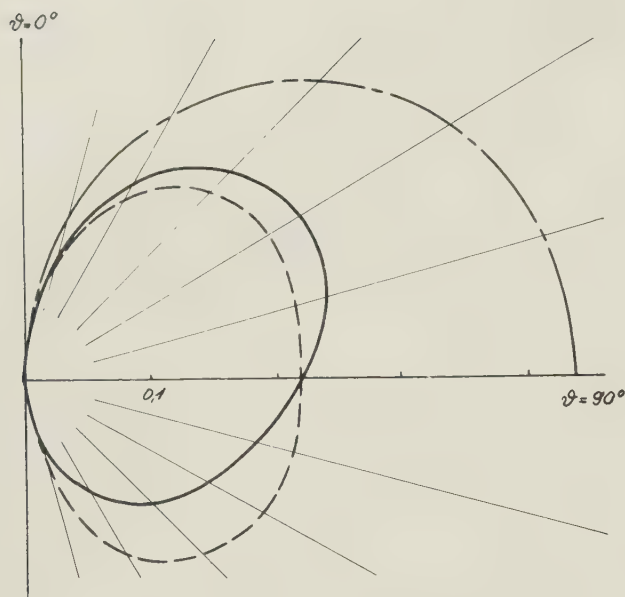
$$H_\varphi \sim - \sqrt{\frac{\mu}{\epsilon}}\, Ue^{-ikr}(k\varrho_o/r)J_1(k\varrho_o \sin \vartheta)(z > 0) \tag{6}$$
$$= 0 \quad (z < 0).$$

μ is the magnetic permeability, ϵ the dielectric constant. The field is solenoidal, so the other components of $\overrightarrow{H}$ vanish. The current distribution induced on the plane of the antenna can be easily computed.

The current has only a radial component; it can be expressed as

$$j(r) = -H_\varphi \quad (\vartheta = 0), \tag{7}$$

and at a sufficiently great distance from the slit the asymptotic expression (6) can be inserted into (7).

The second antenna has been extensively dealt with and we may refer to that work for details.[2,3]

Now we cut off the current distribution outside $\varrho = \varrho_1$, thus constructing the corresponding antenna of the third type. With the method of the vector potential one gets for a radial current distribution $j_1(r)$ the magnetic field

$$H_\varphi^* = \mp \frac{k}{2} \cos \vartheta \frac{e^{-ikr}}{r} \int_0^\infty j_1(\varrho)J_1(k\varrho \sin \vartheta)\varrho\, d\varrho$$
$$(z \gtrless 0).$$

For $j_1(\varrho)$ we choose $-j(\varrho)$ from (7) in the cutoff region $\varrho > \varrho_1$ and 0 for $\varrho < \varrho_1$, and obtain, inserting (6) with $\vartheta = \pi/2$,

$$j_1(\varrho) = (\mu/\epsilon)^{1/2}U \cdot (e^{-ik\varrho}/\varrho)k\varrho_o J_1(k\varrho_o) \quad (\varrho > \varrho_1) \tag{8}$$
$$= 0 \quad (\varrho < \varrho_1).$$

Combining the fields H_φ and H_φ^* we finally arrive at the asymptotic results for large r:

$$\Phi = -k\varrho_o J_1(k\varrho_o \sin \vartheta) - \tfrac{1}{2}k^2(\varrho_o/r)$$
$$\cos \vartheta \cdot J_1(k\varrho_o) \int_{\varrho_1}^\infty J_1(k\varrho \sin \vartheta)e^{-ik\varrho}\, d\varrho \quad (z \geqq 0), \tag{9a}$$

$$\Phi = \tfrac{1}{2}k^2(\varrho_o/r) \cos \vartheta \cdot J_1(k\varrho_o) \int_{\varrho_1}^\infty J_1(k\varrho \sin \vartheta)e^{-ik\varrho}\, d\varrho$$
$$(z \leqq 0), \tag{9b}$$

with the abbreviation

$$\Phi = (\epsilon/\mu)^{1/2}re^{ikr}H_\varphi/U. \tag{10}$$

Numerical Results

In the figures we give some numerical results for several values of ϱ_o, ϱ_1 and wave number k. They refer to the following choices:

1. $k\varrho_o = 1$, $k\varrho_1 = 2$,
2. $k\varrho_o = 1$, $k\varrho_1 = 4$,
3. $k\varrho_o = 10 \sin 5^0 = 0, 87$, $k\varrho_1 = 10$,
4. $k\varrho_o = 10 \sin 10^0 = 1, 74$, $k\varrho_1 = 10$,
5. $k\varrho_o = 10 \sin 15^0 = 2, 59$, $k\varrho_1 = 10$,

and represent the magnitude of Φ, each for the three different antennas. The dash-dotted, the full, and the dotted lines give Φ as a function of the angle ϑ for the first, second and third antenna respectively.

Fig. 1

Fig. 3

Fig. 2

approximation of the radiation pattern. But the radiation pattern, even in the case of Fig. 1, of the cutoff current antenna, is rather near the pattern of the antenna of the second type. Of special interest is the fact that this approximation also gives, at least qualitatively, the back radiation at angles ϑ between 90 degrees and 180 degrees.

Considering Figs. 3 to 5 with the same value $k\varrho_1 = 10$ corresponding to a diameter of the disk of about wavelengths, we find these results confirmed. The diameters of the slit are about $\frac{1}{2}$, $\frac{2}{3}$, 1 wavelength in these figures. Of course the approximation by the cutoff current antenna gets a little poorer with increasing slit diameter but in all cases it is much better than the approximation by the infinite plane antenna.

Analogous results hold for the real and imaginary parts of Φ separately.

Conclusion

Our numerical results show that an antenna of the second type can, as far as the radiation pattern is concerned, be replaced with good approximation by the cutoff current antenna of the third type, even for rather small disks. We must admit however, that there may be a z-component of $\overrightarrow{H}$ for directions in the plane in the approximating radiation pattern which is not present in the actual antenna of the second type. But we do not think that it assumes large values.

One thing we do not obtain by our approximation method is a sufficiently good value for the real and imaginary parts of the impedance of the antenna. This quantity is not only very sensitive to the slit width
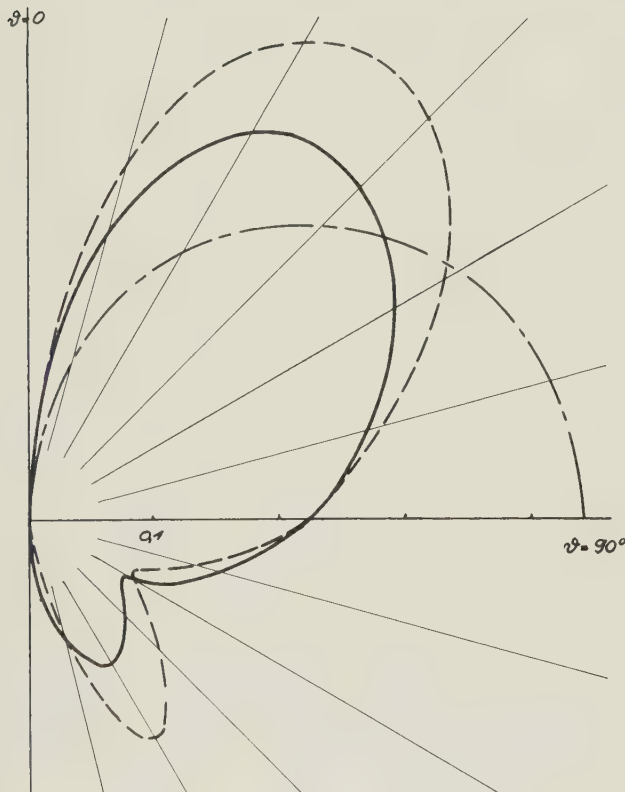
Considering Figs. 1 to 3 with approximately the same value $k\varrho_o$ corresponding to a diameter of the circular slit of about one third of the wavelength, while the outer diameter of the second and third type of antenna is about $\frac{2}{3}$, $\frac{4}{3}$, and 3 wavelengths, we see that in all cases the replacement of the finite circular slit antenna by a circular slit in the infinite plane gives a rather poor
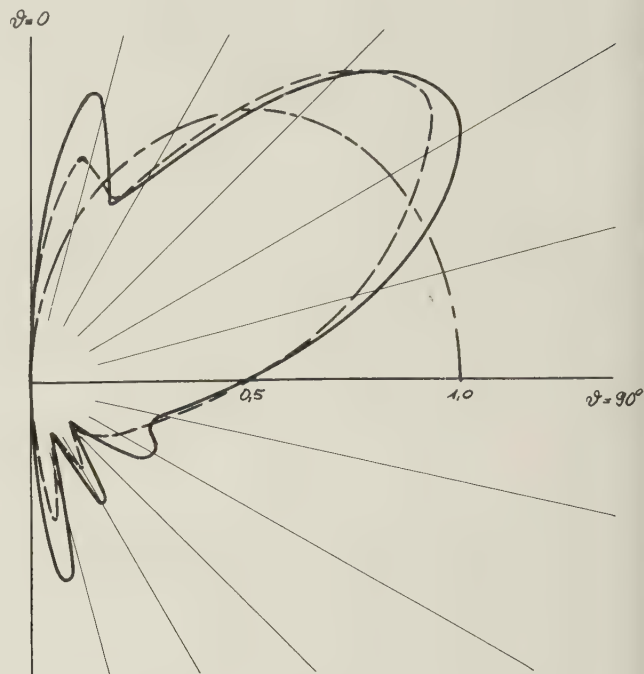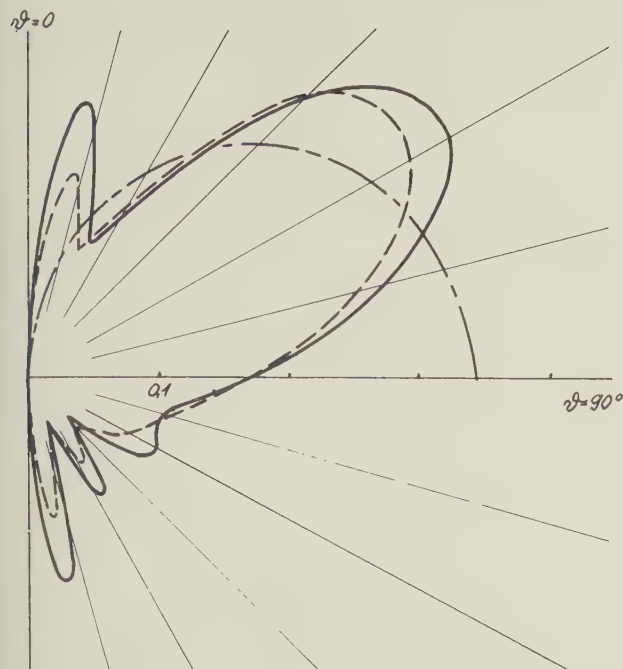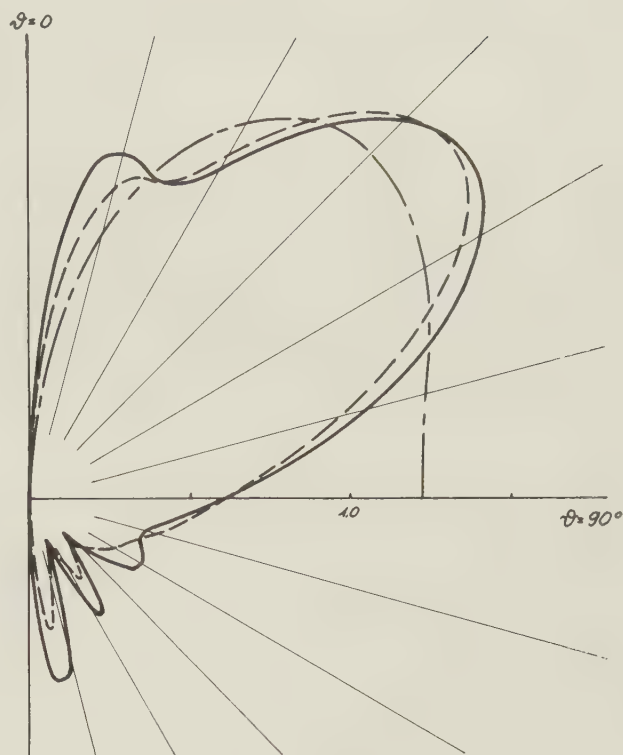
Fig. 4

Fig. 5

but also to small deviations from the correct current distribution near the rims of the slit. By cutting off the current distribution we violate the condition that the current normal to the boundary of the finite antenna vanishes. An adjustment of the cutoff current such that the boundary condition is satisfied, necessarily influences the current distribution essentially over at least some wavelengths and in the cases shown in Figs. 1 to 5 this may involve a noticeable change of the current at the rims of the slit. Of course for large antennas the situation is much more favorable.

Some qualitative indication as to the dependence of the impedance on the size of the antenna may be drawn

from results by Storer[4] on an antenna over a large circular screen.

Acknowledgment

[4] J. E. Storer, "The impedance of an antenna over a large circular screen," *Jour. Appl. Phys.*, vol. 22, pp. 1058–1066; 1951.

Aberrations in Circularly Symmetric Microwave Lenses[*]

M. P. BACHYNSKI[†] AND G. BEKEFI[†]

Summary—The electric field intensity distribution was measured in the image space of solid dielectric microwave lenses at a wavelength of 1.25 cm for various displacements of the source from the principal axis of the system. The experimental results are presented in the form of contours of constant intensity in several receiving planes and also as plots of field intensity versus radial positions of the point of observation. It was found that the deviations of the intensity patterns from the ideal, Airy aberration-free distribution could be interpreted quantitatively in terms of the third order aberrations of optics. The very good agreement obtained with the scalar diffraction theory of aberrations suggests the usefulness of the optical concepts in their application to the centimetric wavelength range.

INTRODUCTION

AS A RESULT of the steady progress that is being made in improving the electrical and mechanical properties of dielectric materials, the lens is beginning to find increasing application at very short radio wavelengths and it may, in the not too distant future, supplant the less versatile conventional metal reflector.

The ideal microwave lens, as is the case with optical lenses, cannot be realized in practice. The deviations of the emergent wave surfaces from the ideal spherical or plane phase fronts cause a deterioration of the image which is as undesirable to the microwave engineer as it is to the designer of optical instruments. In view of the fact that most microwave systems are limited to sizes seldom exceeding about one hundred wavelengths, one can never escape from the large diffraction effects which mutilate the relatively simple geometrical optics aberration figures well known in visible light. Hence, the microwave engineer must turn to a diffraction theory of aberrations if he ever hopes to unravel the complex patterns that are observed.

A systematic theoretical or experimental study of aberrations at microwave frequencies has not been carried out to date. The investigations reported in the past have been of a semi-quantitative nature and generally restricted to the simpler case of systems with cylindrical symmetry.[1-9]

In the course of the work described here, the question arose of whether or not the concepts of aberrations taken over from physical optics were useful and if so, were strictly applicable in the range of centimeter waves. In the study of this problem much help was obtained from the recent advances made by Nijboer,[10] Nienhuis,[11] Zernike,[12] and Maréchal[13] in their work on the diffraction theory of optical aberrations.[14]

The investigations to be described are concerned with solid dielectric, rotationally symmetrical microwave lenses. Employing favorable experimental arrangements and with the aid of specially corrected lenses, systems affected by virtually pure aberrations were examined; the effects on the diffraction patterns of pure spherical aberration, nearly pure coma and pure astigmatism were observed. A form of the Fresnel-Kirchhoff diffraction integral together with the theory of geometrical optical aberrations were used for comparison with measurements. The good quantitative agreement that was obtained led to the conclusion that the optical concepts were applicable and of considerable importance in the field of microwave optics.

Investigations concerned with the problem of a mixture of aberrations present simultaneously, a case encountered in practice, with the measurement of phase and with other related topics, are discussed elsewhere.[15-1]

DIFFRACTION THEORY OF ABERRATIONS[18]

The electromagnetic energy issuing from a small source is assumed to enter the image space via the exit pupil of the optical system only, while the remainder of the plane occupied by the pupil is taken as perfectly impervious to the waves. Hence the exit pupil acts as a diffracting aperture in an infinitely large, plane screen (see Fig. 1). Due to the difficulties that exist in applying electromagnetic theory to the problem at hand, a form of Kirchhoff's theory will be used; from Dirichlet's

[1] R. C. Spencer and P. M. Austen, Mass. Inst. Tech. Rad. Lab. Reps. No. 762–1; 1945, and 762–2; 1946.

[2] J. Ruze, Cambridge Field Station Rep. No. E5043; 1949.

[3] R. A. Woodson, AF Cambridge Res. Center Rep. No. (19–122) 50–4114; 1950.

[4] R. C. Spencer, C. J. Sletten and J. E. Walsh, AF Cambridge Res. Center Rep. No. E5069; 1951.

[5] R. C. Spencer, AF Cambridge Res. Center Rep. No. E5082; 1951.

[6] T. C. Cheston and D. H. Shinn, *Marconi Rev.*, vol. 15, p. 174; 1952.

[7] R. C. Gunter, paper presented at the IRE Convention, Long Beach, California, also paper No. 36 presented at the Symposium on Microwave Optics, McGill University, Montreal; 1953.

[8] A. S. Dunbar, "Astigmatic diffraction effects in microwave lenses." *Trans.* IRE, vol. 4, pp. 72–80, November, 1952.

[9] K. Milne, Proc. on a Conference on Centimeter Aerials for Marine Navigation, Her Majesty's Stationary Office; 1953.

[10] B. R. A. Nijboer, Thesis, University of Groningen; 1942; *Physica*, vol. 10, p. 679; 1943, vol. 13, p. 605; 1947.

[11] K. Nienhuis, Thesis, University of Groningen; 1948; K. Nienhuis and B. R. A. Nijboer, *Physica*, vol. 14, p. 590; 1949.

[12] F. Zernike and B. R. A. Nijober, "La théorie des Images Optiques," Éditions de la Revue d'Optique, p. 227; 1949, also "Optical Image Evaluation", *Nat. Bur. Standards*, Circular 526, pp. 1–8; 1954.

[13] A. Maréchal, *Rev. d'Opt.*, vol. 26, p. 257; 1947, vol. 27, p. 73, 269; 1948.

[14] For a good review of the subject see E. Wolf, *Reports on Progress in Physics*, vol. 14, pp. 95–120; 1951.

[15] M. P. Bachynski, Thesis, McGill University; 1955.

[16] M. P. Bachynski and G. Bekefi, Eaton Electronics Res. Lab. Tech. Rep. No. 35; 1955. To the U. S. AF Cambridge Research Center on Contract AF 19(122)–81. This paper summarizes some of the results described in the Report.

[17] M. P. Bachynski and G. Bekefi, to be presented at the Symposium on Electromagnetic Wave Theory, University of Michigan, June, 1955.

[18] This chapter is intended only as a summary of the more important results, some of which appear to be new. Most of the notation is substantially the same as that used by Nijboer. Final formulas for purposes of calculation will be found in references 10, 11, 12 and some new formulas in references 15 and 16.
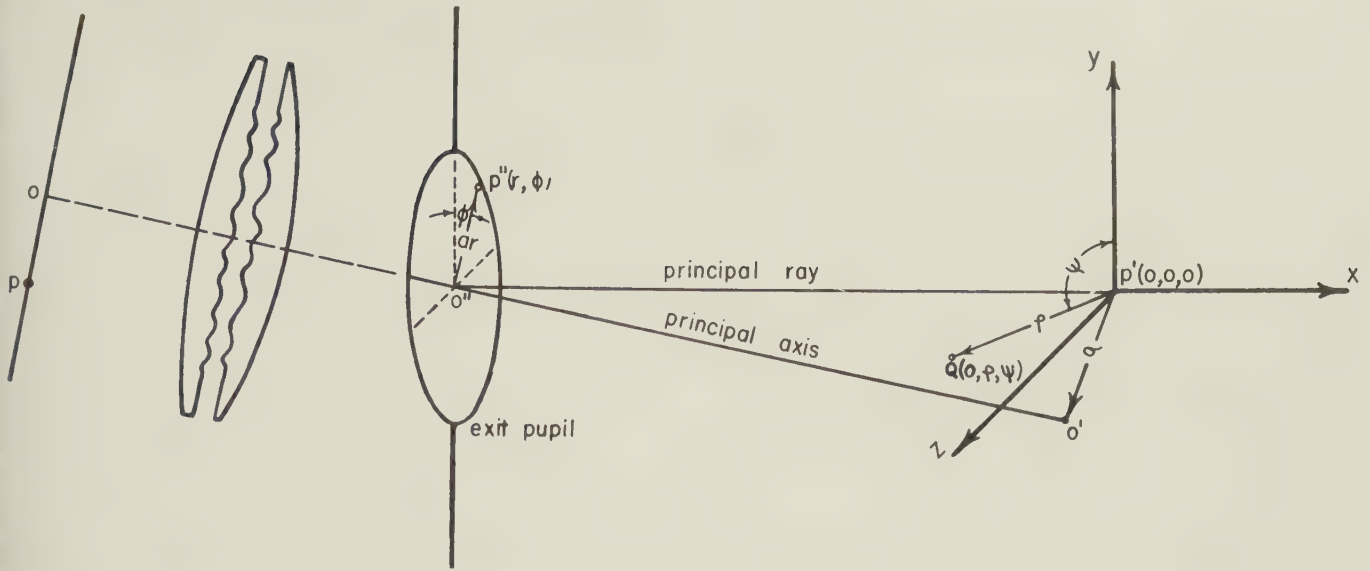
Fig. 1—Coordinate system for the diffraction theory of aberrations. p' is the image point; its conjugate in the object space being p. A right-handed orthogonal system with the origin at the image point and the x-axis coinciding with the principal ray $o''p'$ is used in the image space. $o'p' = \sigma$ represents the distance from the image point to the principal axis of the system. Any point in the image space can be located by either its Cartesian co-ordinates (x, y, z) or its polar co-ordinates (x, φ, ψ). p'', any point in the exit pupil, it described by the polar co-ordinates (r, ϕ) with the origin located as the center o''. The radial coordinate in the exit pupil is normalized so that $0 \le r \le 1$ and the actual distance to any point is given by (ar) where a is the radius of the exit pupil.

boundary value problem[19] and the postulates that the field is zero on the shadow side of the screen and equal to that of the incident wave in the aperture, it follows that the amplitude u_Q at some point Q in the image space is given by

$$u_Q = \frac{1}{2\pi} \int_{S_o} u_i \frac{\partial}{\partial n} \left[\frac{\exp(-iks)}{s} \right] dS, \qquad (1)$$

where the amplitude function u_Q will be associated with that component of the electric field which is oriented in the direction of initial polarization. The frequency dependence $\exp(i\omega t)$ of the monochromatic radiation is assumed implicit in the formula, ω being the angular frequency. S_o denotes an integration over the area of the exit pupil and s the distance from an element δs of the pupil to the field point Q; $(\partial/\partial n)$ signifies a differentiation with respect to the normal to the plane of the pupil, with the normal directed into the image space. u_i is the amplitude of the incident wave over the plane S_o. In the case of a positive optical system, it is the amplitude of a wave converging to a point $(0, 0, 0)$ of the image space (Fig. 1). The exact nature of u_i depends upon the type of source used and upon the properties of the optical system. For an isotropic point radiator and a perfect lens system free from transmission losses and aberrations, u_i represents a spherical wave and is given by

$$u_i = u_o \frac{\exp(ikR_o)}{R_o},$$

where u_o is a constant amplitude term and R_o the distance from the center of curvature of the wave $(0, 0, 0)$ to a point in the pupil. In the presence of aberrations,

the converging wave will no longer have a spherical phase front but will deviate from this surface by an amount V, the aberration function.

Furthermore, the radiator may not be isotropic and the illumination may be tapered towards the edges of the lens, a case generally encountered in practice. Hence

$$u_i = u \frac{\exp[ik(R_o - V)]}{R_o}, \qquad (2)$$

where u_o has been replaced by u, a more general function which enables account to be taken of any amplitude variations over the pupil resulting from anisotropy of the source and any transmission losses through the optical system. For a circularly symmetrical optical system, it is easily shown that, on inserting (2) into (1), expressing the variables in the polar coordinates shown in Fig. 1 and on making the usual approximations $(a/R) << 1$, $(\lambda/R) << 1$ and $(\sigma/R) << 1$, the amplitude at Q is given by

$$u_Q(p, q, \psi) = ika^2 \frac{\exp[ik(R - R')]}{2\pi RR'} \int_o^1 \int_o^{2\pi} u(r)$$

$$\exp[ipr^2 - iqr \cos(\phi - \psi) - ikV(\sigma, r, \phi)] r \, dr \, d\phi, \qquad (3)$$

where a is the radius of the exit pupil and R the distance of the center of curvature of the wave $(0, 0, 0)$ from the center of the pupil; the integration over the aperture has been normalized to unity through a change of variable. $u(r)$ is the amplitude distribution over the aperture and is a constant for an isotropic radiator and a system free from transmission losses; it is assumed to be independent of the azimuthal angle ϕ. The remaining quantities in (3) are defined as follows:

$$p = \frac{ka^2}{2RR'} x, \qquad\qquad q = \frac{ka}{R'} \rho.$$

$$R' = [(R + x)^2 + \rho^2]^{1/2}, \quad k = 2\pi/\lambda,$$

[19] Mathematically it is more satisfactory to develop the theory in this manner rather than start from Kirchoff's formula, which contains mixed boundary requirements (see C. J. Bouwkamp, *Reports on Progress in Physics*, vol. 17, p. 35; 1954.

with x denoting the displacement of the receiving plane from the Gaussian image plane along the principal ray of the system.

Henceforth, p instead of x will be used in representing the displacement of the receiving plane and q rather than ρ will be used in describing the radial distance of the point of observation from the principal ray.

Eq. (3) in a slightly different form has been discussed at length by Lommel as early as 1885 for the case of an aberration-free system, $V = 0$, and a uniformly illuminated aperture, $u(r) = $ constant. The special case for the field distribution in the geometrical image plane ($p = 0$) yields the well known Airy ring pattern of an ideal system, the intensity distribution of which is proportional to $[J_1(q)/q]^2$ with J_1 a Bessel function of the first kind.

Hitherto the aberration function $V(\sigma, r, \phi)$ has not been defined explicitly and a detailed discussion will be found elsewhere.[10, 20] It suffices to say that it can be expanded in a series of the form

$$V(\sigma, r, \phi) = \sum_{l,m,n} b_{lnm}\sigma^{2l+m} F_{nm}(r) \cos m\phi,$$

where l, m, n are non-negative integers, $n \geqslant m$, $(n - m)$ is even and $F_{nm}(r)$ a certain polynomial in r. The integer m specifies the type of aberration ($m = 0$, spherical aberration, $m = 1$, coma, $m = 2$ astigmatism) while curvature of field and distortion are found to be degenerate cases of spherical aberration and coma respectively.

The evaluation of the integral in (3) containing even a few terms of the complete aberration function is extremely difficult. As a result, most attempts at a solution of the problem have been confined to an aberration function containing only a single term. In fact little progress in the solution of the diffraction integral, except for the case of pure spherical aberration, has been made until the work of Nijboer.[10] We shall therefore limit the present treatment to the case of a single type of aberration only, in which instance the aberration function can be conveniently written as

$$k V(\sigma, r, \phi) = \frac{K_{nm}}{L} F_{nm}(r) \cos m\phi. \quad (4)$$

Here K_{nm} is a numerical constant which determines the magnitude of a particular aberration such that $(K_{nm}/2\pi)$ is the maximum deviation of the converging wave from the ideal spherical surface in units of wavelengths and L is the numerical coefficient of the first term of the polynomial $F_{nm}(r)$.

A substitution of (4) into (3) leads, after some manipulation, to the following result:

$$u_Q(p, q, \psi) = ika^2 \frac{\exp [ik(R - R')]}{RR'} \sum_{l=0}^{\infty} \epsilon_l i^{l(m-1)}$$

$$\cos (mt\psi)I_{mt}(p, q, K_{nm}); \quad (5a)$$

where

$$\epsilon_t = 1, t = 0$$
$$= 2, t > 0$$

and

$$I_{mt}(p, q, K_{nm})$$
$$= \int_0^i u(r) \exp (ipr^2)J_{mt}(qr)J_t\left[\frac{K_{nm}}{L}F_{nm}(r)\right] r \, dr. \quad (5b)$$

J_{mt} and J_t are Bessel functions of order (mt) and t respectively. When $m = 0$ (spherical aberration) the above formula simplifies to

$$u_Q(p, q, \psi) = ika^2 \frac{\exp [ik(R - R')]}{RR'} \int_0^i u(r)$$

$$\exp \left[ipr^2 - i\frac{K_{no}}{L}F_{no}(r)\right] J_o(qr)r \, dr. \quad (5c)$$

Eqs. (5a) and (5b) appear to be new; the special case (5c) has been derived already (using different notation) by Steward.[21] The expressions given above are general in that they are applicable whatever the magnitude of the aberration coefficient K_{nm} may be. Nijboer's results useful only when the aberrations are not too large (maximum deviation from the ideal sphere of the order or less than one wavelength, i.e., $K_{nm} \leq 6$), are obtained from (5b) on expanding the Bessel function J_t in ascending powers of K_{nm}. Eq. (5b) then becomes

$$I_{mt}(p, q, K_{nm}) = \sum_{l=0}^{\infty} \frac{i^{2l}}{2^{2l+t}l!(t + l)!}\left(\frac{K_{nm}}{L}\right)^{2l+t}$$

$$\int_0^i u(r) \exp (ipr^2)[F_{nm}(r)]^{2l+t}J_{mt}(qr)r \, dr. \quad (6)$$

The significant contribution of Nijboer to the diffraction theory of aberrations lies in that a circle polynomial $R_n^m(r)$, first introduced by Zernike, which is orthogonal within a unit circle, takes the place of the function $F_n^m(r)$. Not only does the use of this polynomial aid greatly the evaluation of the integral in (6), but it is useful in the balancing of higher-order aberrations and results in a more elegant classification of image errors than that used hitherto.

Before quantitative comparisons can be made between computations from (5) and (6) and the experimental results, numerical values must be assigned to the coefficients K_{nm} for each of the aberrations considered. The magnitudes of these coefficients depend upon the optical system and the position of the source. For a thin lens, having spherical or nearly spherical surfaces, and an exit pupil coincident with its plane (criteria fairly closely satisfied in the experiments), the magnitudes of the third order aberration coefficients can be computed from these formulas[20]:

(a) Spherical Aberration:

$$K_{40} = \frac{1}{32} kK^3\left[\frac{\mu + 2}{\mu(\mu - 1)} \left(X - \frac{2 (\mu^2 - 1)}{\mu + 2} Y\right)^2\right.$$

$$\left. + \left(\frac{\mu^2}{(\mu - 1)^2} - \frac{\mu}{(\mu + 2)} Y^2\right)\right]a^4,$$

[20] H. H. Hopkins, "Wave Theory of Aberrations," Oxford University Press, New York, 1950.

[21] C. G. Steward, *Phil. Trans. Roy. Soc.*, A, vol. 225, pp. 131, 1925.

where

$$k = 2\pi/\lambda; \lambda = \text{wavelength}$$

a = radius of exit pupil

K = 1/focal length = power of lens

μ = refractive index of the lens material

$$X = \frac{r_2 + r_1}{r_2 - r_1} = \text{shape factor}$$

r_2, r_1 = radii of curvature of 1st and 2nd lens surfaces respectively

$$Y = \frac{s - s'}{s + s'} = \text{position factor}$$

s, s' are the object and image distances respectively.

(b) Coma:

$$K_{31} = \frac{kK^2}{4} \sin \theta \left[\frac{\mu + 1}{\mu(\mu - 1)} X - \frac{2\mu + 1}{\mu} Y \right] a^3.$$

(c) Astigmatism:

$$K_{22} = \left[k \frac{K}{4} \sin^2 \theta \right] a^2.$$

(d) Curvature of Field:

$$K_{20} = \left[k \frac{K}{4} \left(1 + \frac{1}{\mu} \right) \sin^2 \theta \right] a^2.$$

From the above, the coefficients of the various aberrations can be found as function of the displacement (θ) of the source from the principal axis of the system. (It will be noted that spherical aberration is independent of the angle θ).

The experimental arrangement followed closely the requirements of the theory presented above in order that direct comparisons between calculated and measured values could be made. The lens was mounted in a large metal screen which acted as the aperture stop and with which the exit pupil was found to be nearly coincident. The lens was illuminated uniformly with energy issuing from a small horn 2 by $2\frac{1}{2}$ wavelengths in size. The various aberrations were obtained by a rotation of the lens (through an angle θ) relative to the source. With $\theta = 0$, only spherical aberration was present. One of the lenses, due to the different radii of curvature of its two surfaces was used primarily in the study of coma. A second lens designed to be free of spherical aberration and coma for a particular object distance was employed in the study of astigmatism.

A piece of open waveguide mounted upon an optical bench was used as a field probe for measurements of the field intensity distributions in the image plane of the lens. The signal received from it was rectified and recorded. The motion of the recorder paper was synchronized with that of the mechanically driven probe and in this manner a record of field intensity as a function of the probe position was obtained. A number of precautions and preliminary investigations were carried out to ensure that the results obtained were due solely to the lens aberrations.

Aberrations in Microwave Lenses

Comparison of Theory with Experiment

The investigations described here are concerned mainly with measurements of the intensity distribution in planes lying perpendicular to the principal ray (Fig. 1) for various angular displacements (θ) of the source from the principal axis. Such displacements illustrate the nature of the diffraction pattern as a function of the various aberrations.

The experimental and theoretical data are depicted diagrammatically in two ways: for quantitative comparisons with theory, scans of the electric field intensity distribution along an axis lying in one of these planes and passing through the principal ray are given. Due to the great similarity between scans along the electric and magnetic diameters, only one of these is presented. The second method of showing the results is by means of contours of constant intensity in given receiving planes and these enable one to obtain a complete picture of the distributions.

The majority of the studies were restricted to measurements in certain planes of special interest. In Table I are listed the various planes, together with the form

Table I.

Aberration	Plane of investigation	Term of aberration function $V(\sigma, r, \phi)$	Shift of receiving plane from Gaussian plane in units of p	Remarks
Spherical aberration	Paraxial plane	$K_{40}r^4/k$		
	Plane midway between paraxial and marginal foci	$\frac{K_{40}(6r^4 - 6r^2}{6} + 1)/k$	$-K_{40}$	Plane of "best focus;" pattern approaches closest to ideal.
Coma	Gaussian	$K_{31}r^3 \cos \phi/k$		Lateral shift of diffraction pattern neglected.
	Gaussian	$\frac{K_{31}(3r^2 - 2r)}{3} \cos \phi/k$		Lateral shift included.
Astigmatism	Central plane	$K_{22}r^2 \cos 2\phi/k$	$-K_{22}$	Plane containing circle of least confusion.
	Plane through focal lines	$K_{22}(r^2) \cos 2\phi/k$	0; $-2K_{22}$	

of the aberration function to be used in calculations; the positions of these planes relative to the Gaussian image plane are also shown. In all but the last case of Table I, the shifts have been incorporated into the

Fig. 2—Intensity distribution in the image space for a microwave dielectric lens which is virtually aberration free. The intensity patterns in a series of planes perpendicular to and traversing the principal axis as well as the intensity distribution along the principal axis are shown.

formulation of the aberration function in such a way that p can be set equal to zero in (5) and (6); in the last instance, namely astigmatism in the plane through the focal lines, p must be set equal to $\pm K_{22}$. Curvature of field and distortion have not been tabulated as these have not been investigated experimentally for reasons discussed later.

Before proceeding with a description of the effects of aberrations, it may be of interest to show some results obtained with an optical system virtually free of image errors. Fig. 2 shows a series of patterns taken in the transverse plane at various distances from a lens illuminated at normal incidence ($\theta = 0$). The axial field distribution is also shown. Only in a narrow region close to the image plane (at a distance from the lens of about 360 cm) does the pattern approach the ideal Airy distribution. On either side of this region, the minima fill in and the side lobe levels increase. It will also be observed that the intensity at the center of the "ideal" pattern does not coincide with the plane in which the intensity is a maximum, which lies much nearer to the lens. This curious fact, which can be explained theoretically, is due to the relative size of microwave lenses in comparison to the wavelength and would not be apparent at optical frequencies. A complete intensity pattern measured in the image plane is shown enlarged in Fig. 3. The first side lobe is 17.5 decibels below the central maximum, in agreement with theory for an ideal lens.

Spherical Aberration

Spherical Aberration in the Paraxial Plane. In the paraxial image plane the aberration function for the presence of spherical aberration is

$$V(\sigma, r) = K_{40}r^4/k.$$

The intensity patterns are independent of the angle of rotation of the lens and have been evaluated numerically for the amounts of aberration corresponding to $K_{40} =$
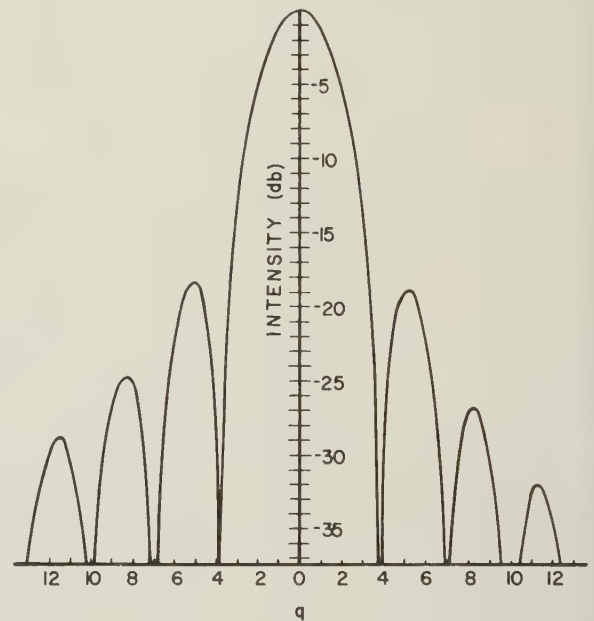

Fig. 3—A nearly ideal intensity distribution in the focal plane of a microwave lens. The first sidelobes are 17.5 decibels down from the main lobe and the first minima about 40 decibels down from the main lobe. The pattern was measured using a microwave dielectric lens corrected for all aberrations at this specific object distance.

$\frac{1}{2}$; 1; 2 and 3 (Fig. 4). It will be observed that as the aberration increases the intensity pattern is altered in that the minima fill in, and the main lobe is decreased while the side lobes rise. With as little spherical aberration as $K_{40} = 3$ (maximum deviation from the reference sphere of 0.48λ) the first few Airy rings have become completely filled in. Fig. 5 depicts a comparison between experiment and theory for nearly equal amounts of aberration. The agreement is seen to be good. No plots of contours of constant intensity are given as these are merely circles and the complete distribution is described by a scan across any diameter.
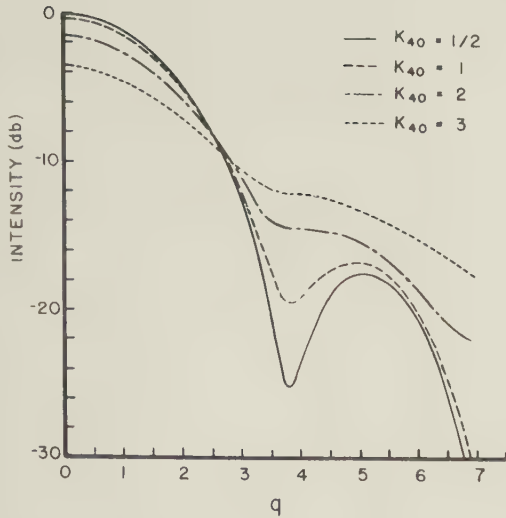
Fig. 4—Intensity distribution in the paraxial focal plane for various amounts of spherical aberration. The results have been normalized to correspond to a value of 0 decibels at the center of the ideal Airy disc.
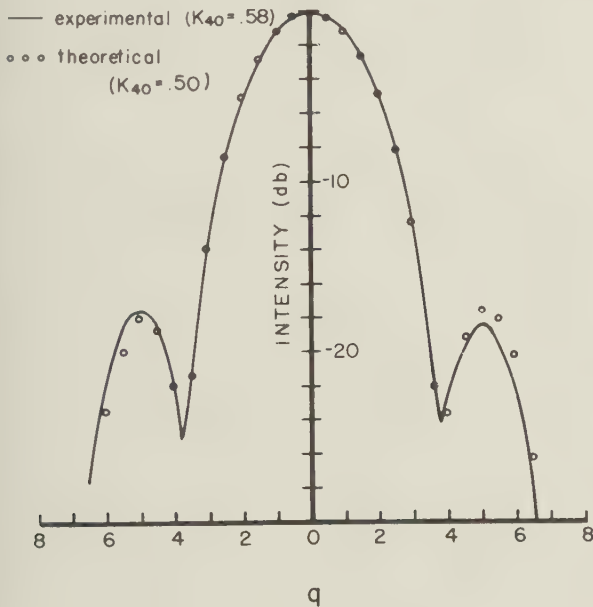

Fig. 5—Intensity distribution in the paraxial focal plane. Experimental amount of spherical aberration corresponds to $K_{40} = 0.58$; theoretical: $K_{40} = 0.5$.
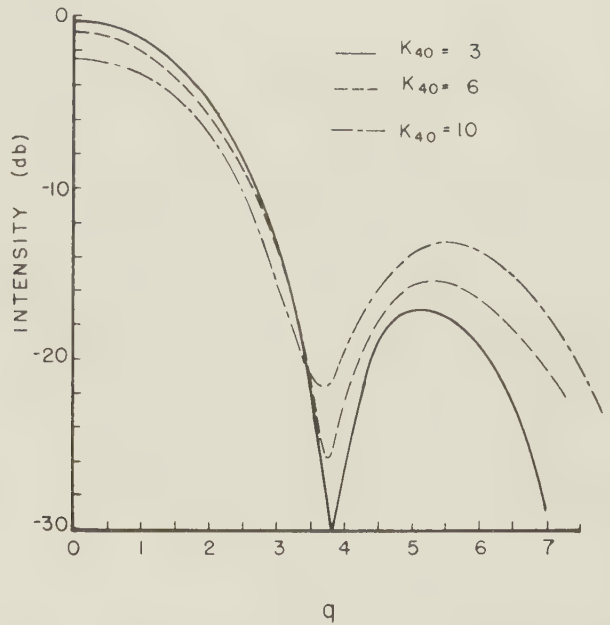

Fig. 6—Intensity distribution in the plane midway between the paraxial and the marginal focus for various amounts of spherical aberration. The intensity at the center of the ideal Airy disc corresponds to 0 decibels.

coefficients of r^4 and r^2 of -1 results in a shift to the position of "best focus." Calculations of the intensity distribution for $K_{40} = 3$, 6 and 10 are presented in Fig. 6. The pattern in this plane varies less rapidly with increasing aberration than in the paraxial plane. No measurements are available for these substantial amounts of aberration.

Coma in the Gaussian Image Plane

The aberration function for the presence of coma may be written either as

$$V(\sigma, r, \phi) = \frac{K_{31}}{3} (R^1_3(r)) \cos \phi/k$$

$$= \frac{K_{31}}{3} (3r^3 - 2r) \cos \phi/k$$

or in the form

$$V(\sigma, r, \phi) = K_{31}r^3 \cos \phi/k.$$

The diffraction pattern is the same whichever of the above expressions is used in the calculations. The additional term $(\frac{2}{3}K_{31}r \cos \phi/k)$ which appears in the first formula causes a shift in the origin of co-ordinates to the position of maximum intensity, for small amounts of coma.[10]

The intensity distribution along the diameter $\psi = 0$ (c.f. Fig. 1) in the Gaussian image plane has been calculated for $K_{31} = 1$; 3 and 9 and is shown in Fig. 7. The coma pattern is far from symmetric about the position $q = 0$, the side lobes on one side gradually become completely suppressed while the opposite side lobes grow in intensity. Coma of an amount as small as $K_{31} = 1$ (maximum deviation of 0.16λ) has greatly altered the appearance of the lobe structure and hence is readily detectable.

Spherical Aberration in the Plane Midway Between the Paraxial and the Marginal Focus. For a system afflicted with spherical aberration, the receiving plane where the resemblance of the intensity distribution is closest to the ideal is located not at the paraxial focus but at a point midway between the paraxial and the marginal focus. This implies a shift of the receiving plane and it can be shown that for small aberrations the expression to be used for the aberration function is

$$V(\sigma, r, \phi) = \frac{K_{40}}{6} R_4^o(r)/k = \frac{K_{40}}{6} (6r^4 - 6r^2 + 1)/k.$$

The r^2 term determines the shift of the receiving plane and it has been shown by Richter[22] that a ratio of the

---

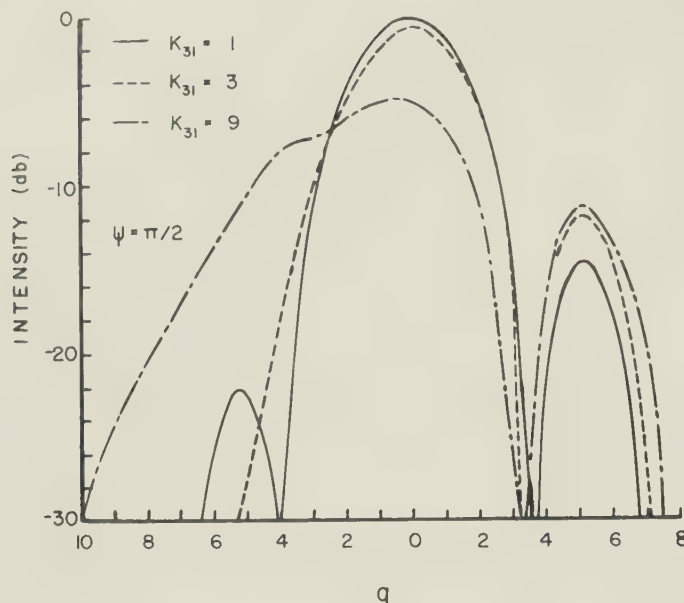[22] R. Richter, *Z. fur. Inst.* vol. 45, p. 1; 1925.

Fig. 7—Intensity distribution in the Gaussian image plane for various amounts of coma.

Contours of constant intensity in the Gaussian image plane are shown in Fig. 8(a). The experimental plots contain predominantly coma ($K_{31} = 1.25$) although a small amount of astigmatism ($K_{22} = 0.37$) which has resulted in some filling in of the minima is also present. These are compared to a set of contours of constant intensity for pure coma of amount $K_{31} = 3$ calculated by Nijboer [Fig. 8(b)]. Although the amount of coma present in the experimental pattern is less than in the theoretical plot, the resemblance is indeed striking. (Nijboer[10] has shown that the general appearance of the coma pattern remains the same for K_{31} values of 1 to 3 or more even though the relative intensities change). Resemblance of the patterns to the geometrical optics coma flare is apparent.

Astigmatism

Astigmatism in the Central Plane. When astigmatism is the only aberration present, the aberration function is given by

$$V(\sigma, r, \phi) = K_{22} r^2 \cos 2\phi / k.$$

Now the plane $p = 0$ lies at the midpoint of astigmatic separation (i.e., the central plane) if $V(\sigma, r, \phi)$ is defined as above.

Calculations of the intensity distribution in the central plane along the diameter $\psi = 0$ or $\psi = \pi/2$, have been made for $K_{22} = \frac{1}{2}$; 1; 2 and 4 and are shown in Fig. 9. The pattern remains symmetrical, the minima fill in rapidly and the central lobe decreases in intensity with increasing values of K_{22}. The side lobes increase with the aberration. However, as the aberration increases (e.g., $K_{22} = 4$) the central lobe is no longer a maximum relative to the first side lobe.

Astigmatism in a Plane through a Focal Line. In order to calculate the intensity distribution in a plane through a focal line, the receiving plane must be shifted in the direction of the positive p-axis by an amount

Fig. 8—(a) Contours of constant intensity measured in the Gaussian image plane of a microwave dielectric lens system in the presence of predominant coma ($K_{31} = 1.25$). A small amount of astigmatism ($K_{22} = 0.37$) is also present. The intensity at the center of the pattern is 890. (b) Contours of constant intensity (theoretical) in the Gaussian image plane in the presence of third-order coma of amount $K_{31} = 3$. (After Nijboer). Intensity at the center is 879.

$$p = \pm K_{22}; \text{ or } x = \pm \frac{2K_{22}R^2}{ka^2 \mp 2K_{22}R}.$$

Experimental Investigations of Astigmatism. Contours of constant intensity in the central plane have been obtained for K_{22} values of $\frac{1}{2}$ and 1. These are illustrated in Fig. 10(a) and Fig. 11(a) (on a following page), where they are compared to calculated contours of constant intensity of the same K_{22} values [Figs. 10(b) and 11(b)] obtained from the work of Nijboer.[10] Good agreement between experiment and theory is observed.

Calculated intensity contours in a plane through a horizontal focal line in the presence of astigmatism of

Fig. 9—Intensity distributions in the central plane for various amounts of astigmatism.

(a)

(b)
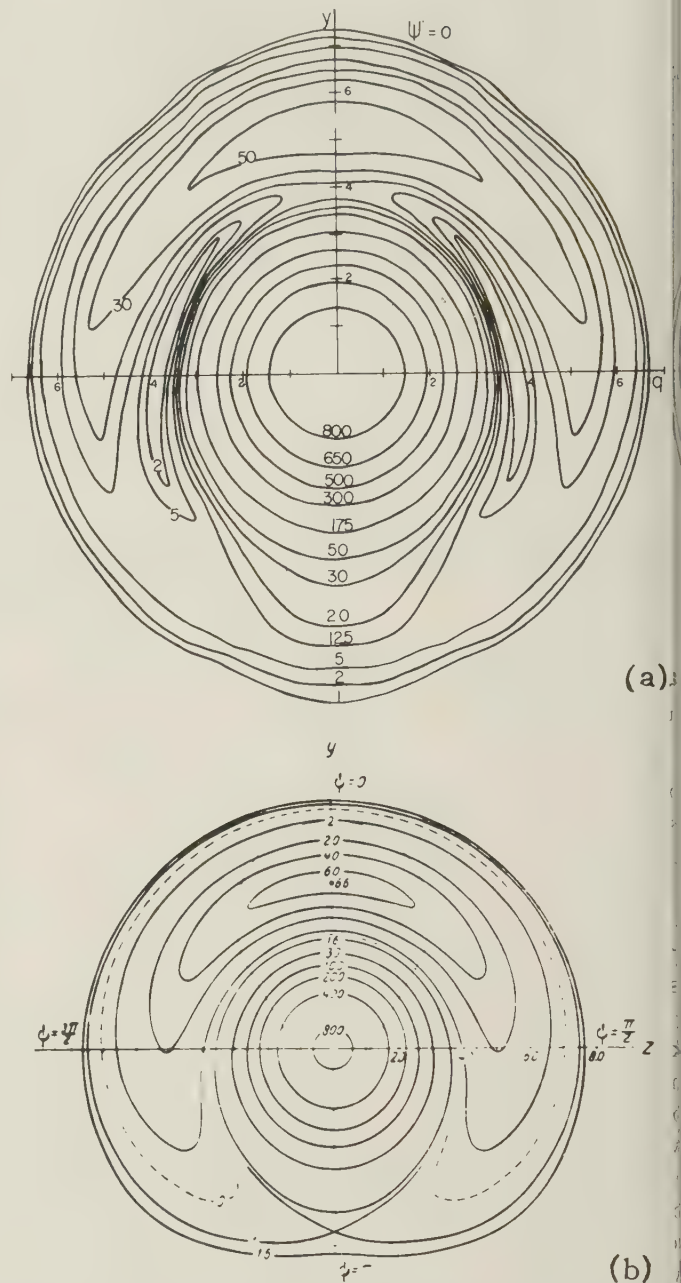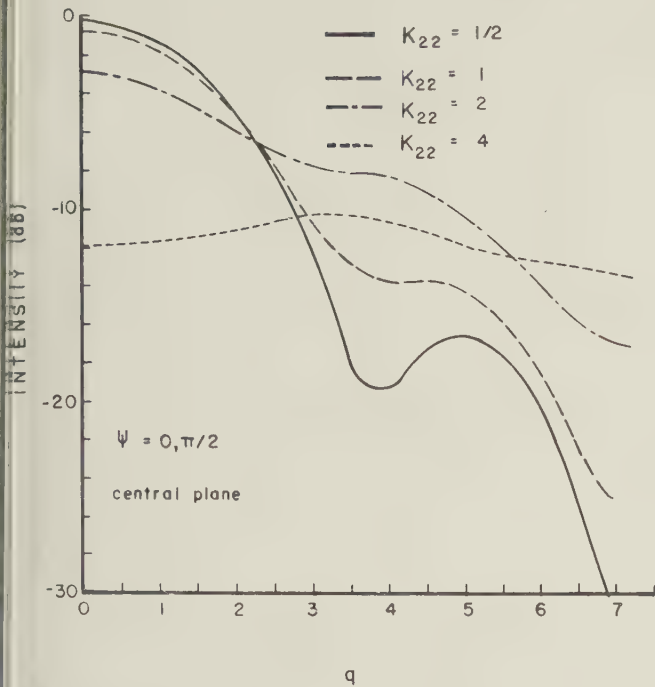
Fig. 10—(a) Contours of constant intensity measured in the central plane of a microwave dielectric lens system in the presence of third-order astigmatism of amount $K_{22} = 1/2$. Intensity at the center of the pattern is 959. (b) Contours of constant intensity (theoretical) in the central plane in the presence of third-order astigmatism of amount $K_{22} = 1/2$ (After Nijboer). Intensity at the center is 959.

mount $K_{22} = 1$ are presented in Fig. 12(a). The intensity distribution along the y-azis does not deviate appreciably from that in the image plane of the ideal pattern. However, in the plane at right angles (along the z-axis) the difference is great, resulting in a slow decrease of intensity with increasing q-values and a substantial elongation of the resulting pattern. The positions of the minima have been displaced towards the z-axis from their corresponding positions in the central plane. The theoretical intensity distribution in the plane through the vertical focal line is identical to the one shown, except that it is rotated through the angle $\pi/2$. The experimental intensity contours in the planes through the horizontal and vertical focal lines for $K_{22} = 1$ are depicted in Figs. 12(b) and 12(c) respectively. They show the same distribution as the theoretical plots. Most noteworthy is the 90-degree crossover in the contours upon going from the horizontal to the vertical focal line.

In Figs. 13(a) and 13(b) are compared the experimental and theoretical intensity distributions in the central plane for $\psi = \pi/4$ for amounts of astigmatism $K_{22} = \frac{1}{2}$ and 1 respectively. For $\psi = \pi/4$, the scan passes through the minima positions and is the "best" pattern in this receiving plane with regard to the magnitude of the main lobe above the side lobes. The side lobes are seen to decrease with increasing astigmatism.

The intensity in the central plane along $\psi = 0$ or $\pi/2$ for $K_{22} = \frac{1}{2}$ is shown in Fig. 14. The pattern is symmetrical but the minima have filled in and the side lobes have increased relative to the ideal pattern. The first minimum is 19.5 db down and the first side lobe maximum 16.5 db down from the central maximum.

In Fig. 15 is represented a typical intensity distribution measured for a "large" value of aberration, namely

K_{22} approximately equal to 5. The pattern demonstrates that at these magnitudes of the image error the main lobe is no longer a maximum but has been reduced to a value less than that of the first side lobes. The lobe structure is no longer apparent and the intensity has been spread out drastically.

Curvature of Field and Distortion

The total curvature of field is given by

$$V(\sigma, r) = K_{20}r^2/k.$$

The aberration function has no angular dependence and can be grouped with the pr^2 term of the diffraction
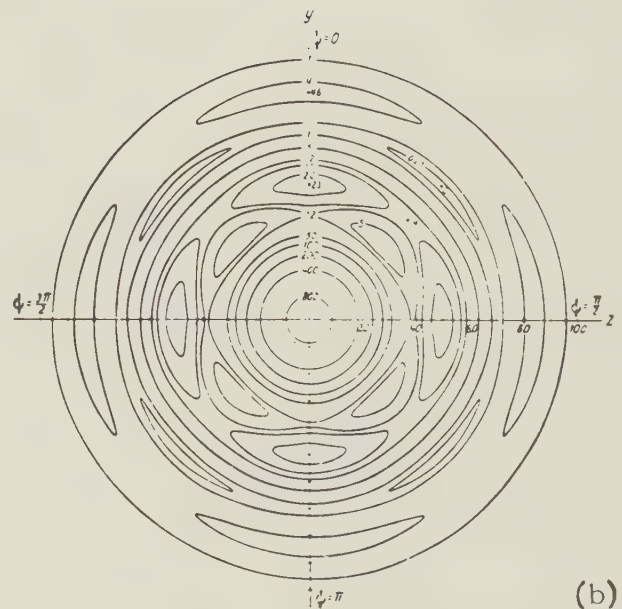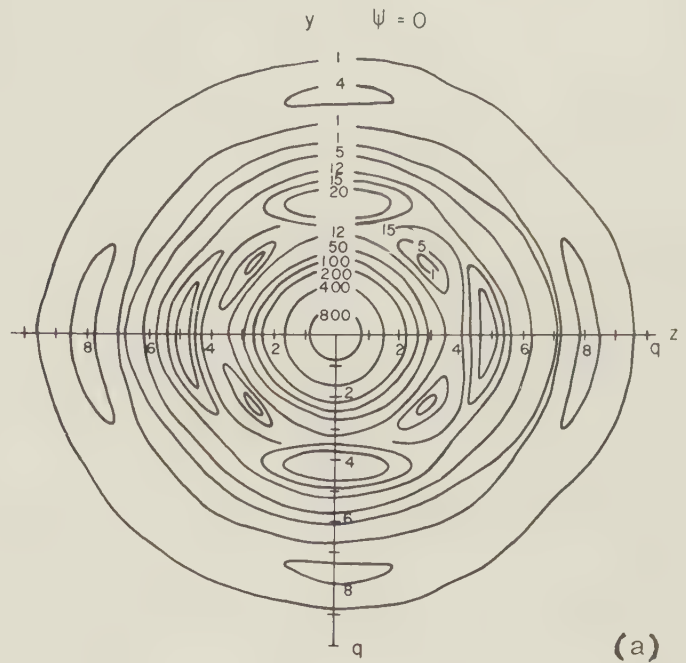
(a)

(b)
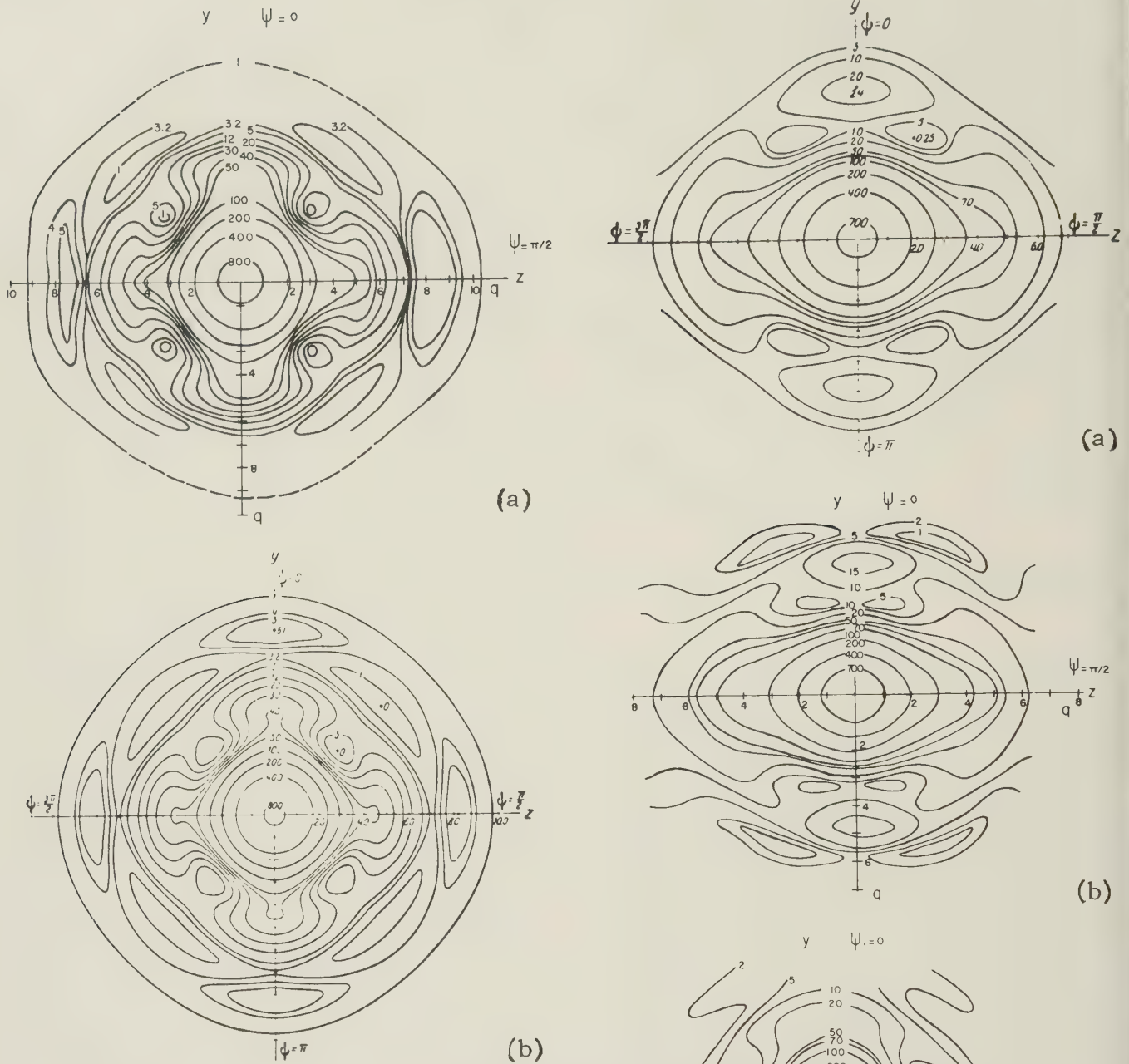
Fig. 11—(a) Contours of constant intensity measured in the central plane of a microwave dielectric lens system in the presence of third-order astigmatism of amount $K_{22} = 1$. Intensity at the center is 840. (b) Contours of constant intensity (theoretical) in the central plane in the presence of third-order astigmatism of amount $K_{22} = 1$. (After Nijboer). Maximum intensity is 840.

integral [see (3)]. This corresponds only to the displacement of the receiving plane along the p-axis. The shift from the Gaussian image plane is given by $p = K_{20}$; i.e.,

$$x = \frac{2K_{20}R^2}{ka^2 - 2K_{20}R}.$$

The resulting diffraction pattern in this new plane is the same as for a spherical wave free of image errors. Since this aberration shows nothing new in the intensity distribution and can be corrected by refocusing (i.e., axial shift of source), it has not been considered separately in experimental investigations.

(a)

(b)

(c)
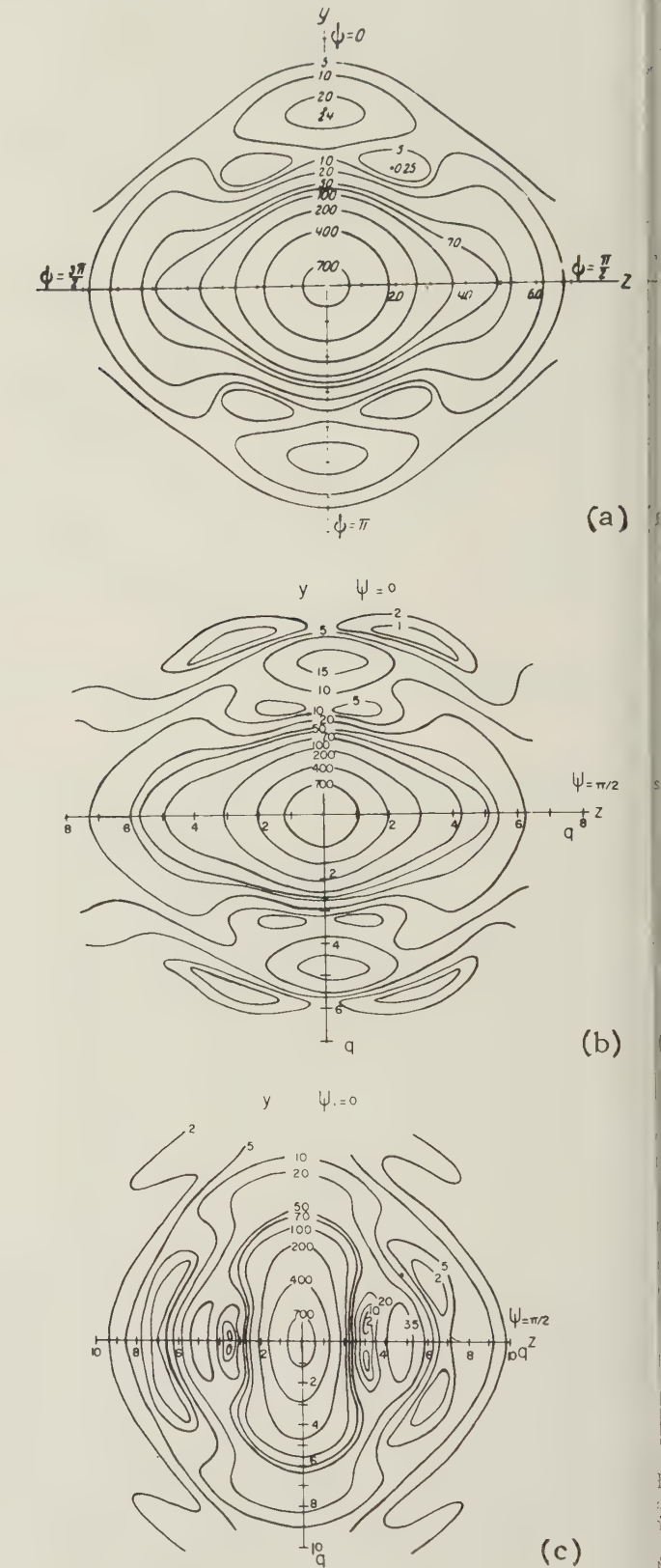
Fig. 12—(a) Contours of constant intensity (theoretical) in the plane through the horizontal focal line in the presence of third-order astigmatism of amount $K_{22} = 1$. (After Nijboer). Maximum intensity is 773. (b) Contours of constant intensity measured in the plane through the horizontal focal line of a microwave dielectric lens system in the presence of third-order astigmatism of amount $K_{22} = 1$. Intensity at the center is 773. (c) Contours of constant intensity measured in the plane through the vertical focal line of a microwave dielectric lens system in the presence of third-order astigmatism of amount $K_{22} = 1$. Intensity at the center is 773.

Fig. 13—(a) Experimental and theoretical intensity distributions in the central plane across the radius $\psi = \pi/4$ in the presence of third-order astigmatism of amount $K_{22} = 1/2$. (b) Experimental and theoretical intensity distributions in the central plane across the radius $\psi = \pi/4$ in the presence of third-order astigmatism of amount $K_{22} = 1$.
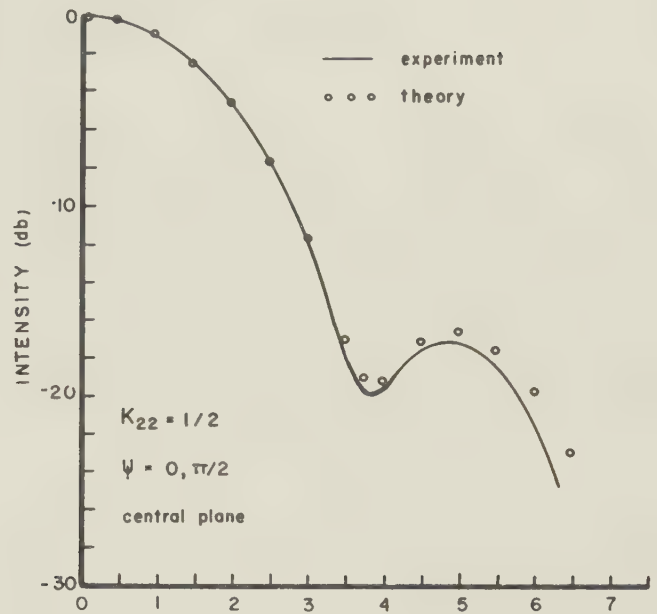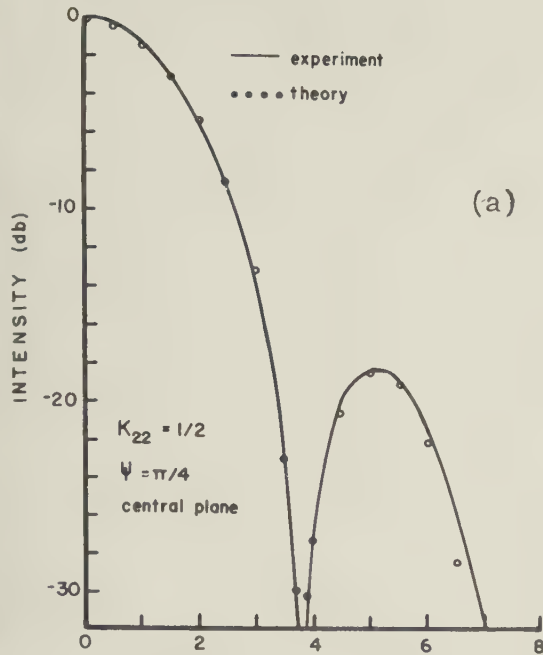
Fig. 14—Experimental and theoretical intensity distribution in the central plane across the radius $\psi = 0$ or $\pi/2$ in the presence of third-order astigmatism of amount $K_{22} = 1/2$.
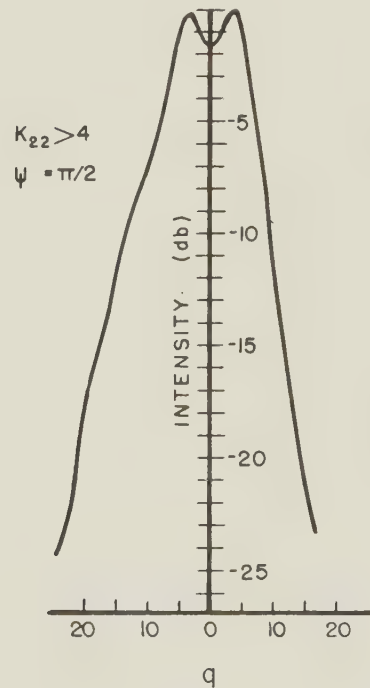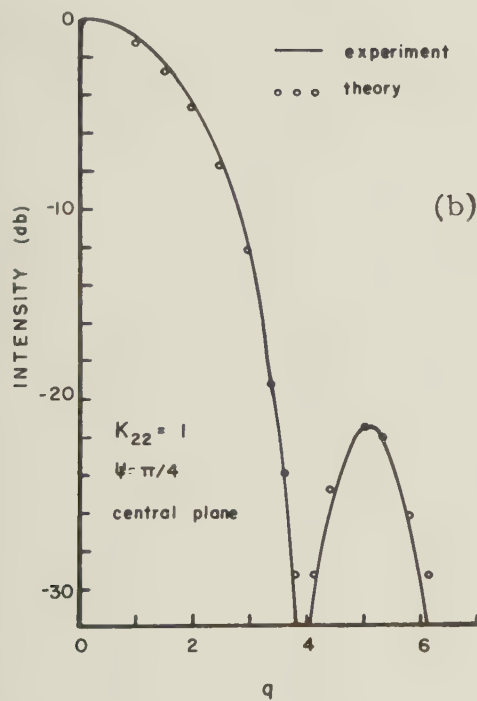
Fig. 15—Intensity distribution in the central plane in the presence of large amounts of astigmatism ($K_{22} > 4$).

Similarly the term for distortion ($V(\sigma, r, \phi) = K_{11}r \cos \phi$) can be grouped with the term ($qr \cos (\phi - \psi)$) of (3), resulting in a different q-value and hence in a lateral shift of the diffraction pattern as a whole.[1]

Spherical Surface-Wave Antennas*

R. S. ELLIOTT†

Summary—Solutions of Maxwell's equations are presented which approximately satisfy the boundary conditions for corrugated and dielectric-clad conducting spheres. These solutions have the physical interpretation of leaky latitudinal surface waves. Values of the complex propagation constant are given as functions of the geometry. For large spheres the leakage is small and the transmission properties approach those of a trapped cylindrical wave on a flat surface.

A corrugated spherical cap, used to support surface waves, has been found to have interesting possibilities as a low-drag omni-directional antenna. Preliminary experimental results are offered as an illustration of the theory.

INTRODUCTION

IN AN EARLIER paper[1] the properties of azimuthal surface waves on circular cylinders were considered. It was found that for uniformly modified cylinders, the surface waves leaked exponentially. This fact was turned to good advantage in terms of pattern control; the cylindrically-curved surface-wave antenna was found to have certain null-filling as well as beam-directing advantages over its plane counterpart.[2] This conclusion is again found to be true in the case of spherical systems.

Beginning in 1951, the Stanford group reported an interesting series of experiments with annular ring antennas.[3-5] By corrugating a flat disk and feeding it from the center, they were able to obtain an endfire-type omnidirectional pattern suitable for beacon applications. The pattern suffered from the defect of deep nulls characteristic of all endfire systems, the main beam was tilted up off the plane of the disk, and the corrugations limited the application to perpendicular polarization. Improved performance was obtained when the corrugations were replaced by a tapered dielectric.

From results obtained by bending flat, linear, corrugated surfaces into cylindrical sections, it was felt that if the Stanford corrugated disk were replaced by a corrugated spherical cap, some filling-in of the nulls could be achieved together with a decrease in the beam tilt angle. This has proved to be the case. If a dielectric film were substituted for the corrugations, the spherical cap should be capable of supporting either polarization (and consequently, any arbitrary polarization).

As a first step in the investigation of these ideas, the properties of surface waves on spheres were studied

with the aim of providing information from which the spherical cap could be designed. The earlier sections of the present paper report this study and include discussions of TM surface waves on corrugated spheres and both TM and TE surface waves on dielectric-clad spheres. The paper concludes with a description of a corrugated spherical cap antenna and a comparison of its experimental performance with that of a corrugated disk antenna.

THE CORRUGATED SPHERE

The first antenna problem to be treated will be TM surface waves on a corrugated sphere. Referring to Fig. 1,
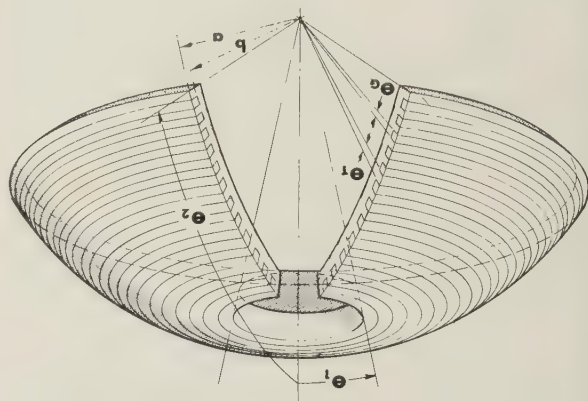
Fig. 1

the waves are assumed to be excited by a ϕ-symmetric source which occupies the region $0 \leqslant \theta \leqslant \theta_1$. They then travel latitudinally over the corrugated surface until the cone $\theta = \theta_2$ is reached. The region $\theta_2 \leqslant \theta \leqslant \pi$ is assumed filled with a perfectly absorbing material so that no reflections occur. (This idealized picture approximates quite well the practical problem of a corrugated spherical cap antenna. In that problem the surface waves, upon reaching the end of the cap, are launched into free space with negligible end reflection.) The desired wave motion therefore has the form

$$e^{-i(wt-\gamma a\theta)},$$

in which a is the radius of the sphere, γ is the propagation constant, and an e^{-iwt} time dependency is assumed. The principal goal of the analysis is to discover if there are circumstances under which γ is principally real, indicating that latitudinal waves essentially nonleaky in character can be made to propagate.

Let Region I ($\theta_1 \leqslant \theta \leqslant \theta_2$, $r \geqslant a$) be taken as free space with dielectric constant ϵ_o and permeability constant μ_o. If MKS rationalized units are used, and TM excitation (TM in this connection means $H_\theta \equiv 0$) assumed, the magnetic field in Region I satisfies the differential equation

† Hughes Aircraft Co., Culver City, Calif.
[1] R. S. Elliott, "Azimuthal surface waves on circular cylinders," *Jour. Appl. Phys.*, vol. 26, pp. 368–376; April, 1955.
[2] R. S. Elliott, "Pattern Shaping with Surface Wave Antennas," Hughes Res. and Dev. Labs., Tech. Memo. No. 360; April, 1955.
[3] "Ridge and Corrugated Antenna Studies," Stanford Res. Inst., Final Report; January, 1951.
[4] E. M. T. Jones, "An annular corrugated-surface antenna," *Proc. IRE*, vol. 40, pp. 721–725; June, 1952.
[5] E. M. T. Jones and R. A. Folsom, Jr., "A note on the circular dielectric-disk antenna," *Proc. IRE*, vol. 41, p. 798; June, 1953.

$$0 = \frac{\partial^2 H_\phi}{\partial r^2} + \frac{2}{r}\frac{\partial H_\phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2 H_\phi}{\partial \theta^2}$$

$$+ \frac{\cot\theta}{r^2}\frac{\partial H_\phi}{\partial \theta} + \left(k_o^2 - \frac{1}{r^2\sin^2\theta}\right)H_\phi, \quad (2)$$

in which $k_o = \omega\sqrt{\mu_o\epsilon_o}$. The electric field is related to the magnetic field by the equations

$$-i\omega\epsilon_o E_r = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta H_\phi) \quad (3)$$

$$i\omega\epsilon_o E_\theta = \frac{1}{r}\frac{\partial}{\partial r}(rH_\phi). \quad (4)$$

The appropriate general solution for (2) is[6]

$$H_\phi^e = h_v^{(1)}(k_o r)[A_v P_v^1 \cos\theta + B_v Q_v^1 \cos\theta]e^{-iwt}, \quad (5)$$

in which v is a separation constant, $h_v^{(1)}$ is a spherical Hankel function of the first kind, P_v^1 and Q_v^1 are solutions of the associated Legendre equation and A_v and B_v are constants. The superscript e suggests that Region I is the external region.

A problem arises in the choice of values for the parameter v. It would be desirable to find a complete sequence of eigenvalues and then to express H_ϕ^e as an orthogonal series of terms similar to (5). This task is made difficult by several factors. Firstly, Floquet's Theorem, which so neatly provided the eigenvalues in rectangular[7] and cylindrical[1] geometries, does not apply here. Secondly, the θ interval is restricted, extending from $\theta_1 > 0$ to $\theta_2 < \pi$. A third complication arises when it is recognized that latitudinal surface waves must leak [this point will be brought out more fully in the discussion following (12)] and therefore that applicable values of v must be complex. Attempting to surmount these difficulties appears unwarranted for the purposes of the present analysis and hence an approximation is introduced at this point. It is *assumed* that the principal part of the magnetic field is a single mode for which $v = \gamma a$; i.e.,

$$H_\phi^e = h_{\gamma a}^{(1)}(k_o r)[A P_{\gamma a}^1 (\cos\theta) + BQ_{\gamma a}^1 (\cos\theta)]e^{-iwt}, \quad (6)$$

in which γa is a complex constant to be determined from the boundary conditions. The ensuing analysis will reveal under what conditions (6) is a good assumption.

Restricting attention to spheres for which $k_o a$ is large, the asymptotic expansions

$$P_{\gamma a}^1 (\cos\theta) = \frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\left(\frac{2}{\pi\sin\theta}\right)^{1/2}$$
$$\left\{\cos\left[(\gamma a + \tfrac{1}{2})\theta + \frac{\pi}{4}\right] + 0\left(\frac{1}{\gamma a}\right)\right\} \quad (7)$$

and

$$Q_{\gamma a}^1 (\cos\theta) = \frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\left(\frac{\pi}{2\sin\theta}\right)^{1/2}\left\{\cos\left[(\gamma a + \tfrac{1}{2})\theta\right.\right.$$
$$\left.\left. + \frac{3\pi}{4}\right] + 0\left(\frac{1}{\gamma a}\right)\right\} \quad (8)$$

$$\left(\delta \leqslant \theta \leqslant \pi - \delta, \delta > 0, |\gamma a| \gg \frac{1}{\delta}\right)$$

[6] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., New York, pp. 399–406; 1941.
[7] R. S. Elliott, "On the theory of corrugated plane surfaces," Trans. IRE, vol. AP-2, pp. 71–81; April, 1954.

are applicable.[8] Neglecting the higher-order terms, if A and B are adjusted to a new constant C so that

$$C = A\sqrt{\frac{2}{\pi}} = iB\sqrt{\frac{\pi}{2}}, \quad (9)$$

then (6) becomes

$$H_\phi^e = h_{\gamma a}^{(1)}(k_o r)\frac{\Gamma(va + 2)}{\Gamma(va + \frac{3}{2})}$$
$$\frac{C}{(\sin\theta)^{1/2}}e^{-i\left\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\right\}}, \quad (10)$$

and the desired latitudinal wave motion is properly displayed. The presence of the factor $(\sin\theta)^{-1/2}$ can be explained by realizing that the power density of the wave thins out as the equator is approached.

Using (3) and (4), the corresponding electric field is found to be

$$-iw\epsilon_o E_r^e = \frac{1}{r}h_{\gamma a}^{(1)}(k_o r)\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\frac{C}{(\sin\theta)^{1/2}}$$

$$\left\{i(\gamma a + \tfrac{1}{2}) + \tfrac{1}{2}\cot\theta\right\}e^{-i\left\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\right\}} \quad (11)$$

$$iw\epsilon_o E_\theta^e = \left\{k_o h_{\gamma a}^{(1)'}(k_o r) + \frac{1}{r}h_{\gamma a}^{(1)}(k_o r)\right\}$$

$$\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\frac{C}{(\sin\theta)^{1/2}}e^{-i\left\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\right\}}. \quad (12)$$

A study of (10), (11), and (12) reveals why γa cannot be entirely real. If it were, the factor $e^{i\gamma a\theta}$ would indicate no radial flow of power as the wave progressed latitudinally, whereas the factor $h_{\gamma a}^{(1)}(k_o r)$ *would* indicate a radial power flow. This contradiction is resolved when γa is complex.

The radial component of impedance for this surface wave at the boundary $r = a$ is

$$Z_r^e = \frac{E_\theta^e}{H_\phi^e} = \frac{h_{\gamma a}^{(1)}(k_o a) + k_o a h_{\gamma a}^{(1)'}(k_o a)}{iw\epsilon_o a h_{\gamma a}^{(1)}(k_o a)}. \quad (13)$$

Turning now to Region II—the region consisting of the spaces between the corrugations—the assumption is made that these spaces are filled with a homogeneous, isotropic dielectric of relative permittivity ϵ/ϵ_o. (This dielectric may be air.) As indicated by Fig. 1, each tooth subtends an angle θ_T and each gap an angle θ_G. The tops of the teeth are at a radius a and the bottoms of the gaps at a radius b. If there are many corrugations per wavelength, the space between any two teeth may be thought of as a shorted biconical transmission line which can only support the TEM mode. The field solution for the mth gap is then

$$H_\phi^i = B_m \frac{\cos k(r - b)}{kr} \quad (14)$$

$$-\frac{iw\epsilon}{k}E_\theta^i = B_m \frac{\sin k(r - b)}{kr}, \quad (15)$$

in which $k = w\sqrt{\mu_o\epsilon}$ and the superscript i suggests that Region II is the internal region. The radial component of impedance for this standing wave, at the boundary $r = a$, and in the mth gap, is given by

[8] Bateman Manuscript Project, "Higher Transcendental Functions," vol. I, p. 162, McGraw-Hill Book Co., New York, 1953.

$$Z_r{}^i = \frac{E_\theta{}^i}{H_\phi{}^i} = -\frac{k}{iw\epsilon}\tan k(a-b). \qquad (16)$$

Since $Z_r{}^i \equiv 0$ over a tooth, the *average* internal radial impedance is

$$Z_r{}^i = -\frac{\theta_G}{\theta_G + \theta_T}\frac{k}{iw\epsilon}\tan k(a-b). \qquad (17)$$

Equating (17) and (13) yields the relation

$$\sqrt{\frac{\epsilon_o}{\epsilon}}\frac{\theta_G}{\theta_G + \theta_T}\tan k(a-b) = -\left\{\frac{h_{\gamma a}{}^{(1)\prime}(k_o a)}{h_{\gamma a}{}^{(1)}(k_o a)} + \frac{1}{k_o a}\right\}, \qquad (18)$$

from which γ can be found in terms of the geometry of the antenna. The ability to find a root γ of (18) is a measure of the validity of assuming a single leaky mode in (10).

Letting $\gamma = \beta + i\alpha$, the interest is in trapped waves; i.e., waves for which $\beta > k_o$ and $\beta \gg \alpha$. When in addition $k_o a$ is large, a first approximation to (18) is

$$\sqrt{\frac{\epsilon_o}{\epsilon}}\frac{\theta_G}{\theta_G + \theta_T}\tan k(a-b) = -\left\{\frac{y'_{\beta a}(k_o a)}{y_{\beta a}(k_o a)} + \frac{1}{k_o a}\right\}, \qquad (19)$$

in which $y_{\beta a}(k_o a)$ is a spherical Neumann function. A plot of (19) is given in Fig. 2. β/k_o is seen to be a slowly varying function of $k_o a$ which smoothly approaches the value for a flat disk as $k_o a$ is increased.
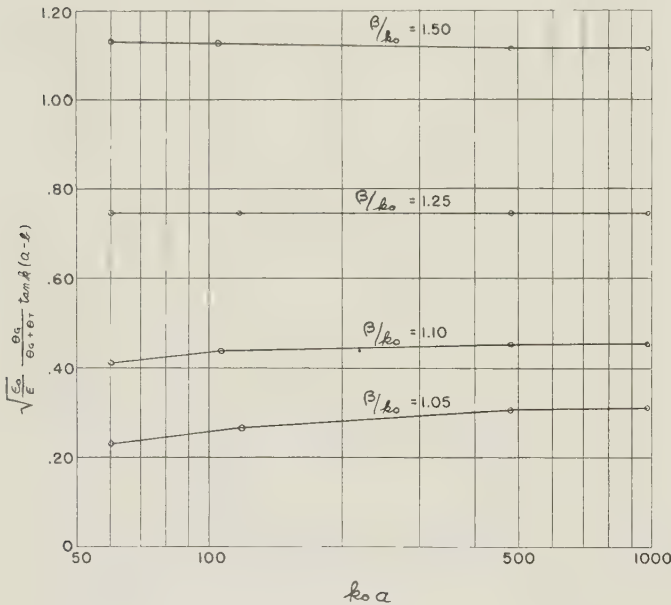
Fig. 2

If now one lets $\gamma a = (\beta a + \delta_1) + i\delta_2$, (18) can be explored for a root by letting βa have the value found from (19) and assuming a succession of values for δ_1 and δ_2. Some representative values of γa are listed in Table I. α (which measures the leakage) is seen to have a considerable, and a useful, dynamic range as a function of $k_o a$ and β/k_o. This feature has been applied to be design of a spherical surface-wave antenna, and will be discussed in the experimental section of this paper.

$k_o a$	$\sqrt{\frac{\epsilon_o}{\epsilon}}\frac{\theta_G}{\theta_G + \theta_T}\tan k(a-b)$	First approximation βa	Second approximation γa
60	.267	62.5	$63 + .9i$
	.419	65.5	$65.5 + .2i$
	.751	74.5	$74.5 + 2 \times 10^{-5}i$
	1.128	89.5	$89.5 + 3 \times 10^{-16}i$
120	.274	125.5	$125.5 + .5i$
	.443	131.5	$131.5 + .01i$
	.751	149.5	$149.5 + 7 \times 10^{-12}i$
	1.123	179.5	$179.5 + 7 \times 10^{-33}i$
480	.311	503.5	$503.5 + .002i$
	.455	527.5	$527.5 + 3 \times 10^{-11}i$
	.750	599.5	$599.5 + 4 \times 10^{-47}i$
	1.119	719.5	$719.5 + 4 \times 10^{-134}i$
960	.316	1007.5	$1007.5 + 10^{-7}i$
	.457	1055.5	$1055.5 + 3 \times 10^{-25}i$
	.750	1199.5	$1199.5 + 2 \times 10^{-95}i$
	1.119	1439.5	$1439.5 + 10^{-266}i$

THE DIELECTRIC-CLAD SPHERE

The geometry of this antenna is suggested by Fig. 3. Once again, a ϕ-symmetric source is assumed to exist in the region $0 \leqslant \theta \leqslant \theta_1$, and to launch surface waves which propagate latitudinally over the sphere. Upon reaching the cone $\theta = \theta_2$, these waves are absorbed by a reflectionless material which fills the region $\theta_2 \leqslant \theta \leqslant \pi$.
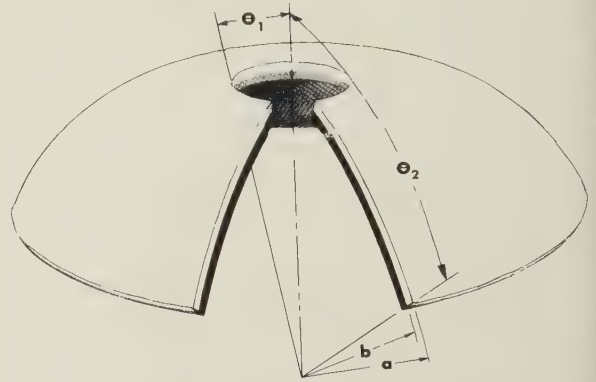
Fig. 3

The volume $\theta_1 \leqslant \theta \leqslant \theta_2$ is composed of three parts. Region I consists of free space with constants μ_o and ϵ_o, and has the range $a \leqslant r < \infty$. Region II consists of a homogeneous, isotropic dielectric with constants μ_o and ϵ and has the range $b \leqslant r \leqslant a$. Region III is a perfect conductor and occupies the space $0 \leqslant r \leqslant b$.

Borrowing from the previous analysis, if TM excitation is assumed, the fields in Region I can be approximated by a single mode of (10), (11), and (12), giving once again the radial impedance at $r = a$ of (13). The fields in Region II are given by

$$H_\phi{}^i = \left\{A j_{\gamma a}(kr) + B y_{\gamma a}(kr)\right\}$$
$$\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}(\sin\theta)^{-1/2}e^{-i\left\{wt - ([\gamma a + \frac{1}{2}]\theta + \pi/4)\right\}} \qquad (20)$$

$$-iw\epsilon E_r^{\ i} = \frac{1}{r}\left\{Aj_{\gamma a}(kr) + By_{\gamma a}(kr)\right\}\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}$$
$$(\sin\theta)^{-1/2}\left\{i(\gamma a + \tfrac{1}{2}) + \tfrac{1}{2}\cot\theta\right\}e^{-i\{\cdots\}} \quad (21)$$

$$iw\epsilon E_\theta^{\ i} = \left\{k[Aj_{\gamma a}'(kr) + By_{\gamma a}'(kr)] + \frac{1}{r}[Aj_{\gamma a}(kr)\right.$$
$$\left. + By_{\gamma a}(kr)]\right\}\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}(\sin\theta)^{-1/2}e^{-i\{\cdots\}}, \quad (22)$$

in which $j_{\gamma a}$ and $y_{\gamma a}$ are spherical Bessel and Neumann functions and $k = w\sqrt{\mu_o\epsilon}$.

Since $E_\theta^{\ i} \equiv 0$ at $r = b$,

$$B = -A\frac{j_{\gamma a}(kb) + kbj_{\gamma a}'(kb)}{y_{\gamma a}(kb) + kby_{\gamma a}'(kb)}, \quad (23)$$

and the internal radial impedance at $r = a$ becomes

$$Z_r^{\ i} = \frac{E_\theta^{\ i}}{H_\phi^{\ i}} = \frac{1}{iw\epsilon a}\left\{\frac{kaj_{\gamma a}'(ka)[y_{\gamma a}(kb) + kby_{\gamma a}'(kb)] - kay_{\gamma a}'(ka)[j_{\gamma a}(kb) + kbj_{\gamma a}'(kb)]}{j_{\gamma a}(ka)[y_{\gamma a}(kb) + kby_{\gamma a}'(kb)] - y_{\gamma a}(ka)[j_{\gamma a}(kb) + kbj_{\gamma a}'(kb)]} + 1\right\} \quad (24)$$

Upon equating (24) and (13), one obtains

$$\frac{j_{\gamma a}(kb) + kbj_{\gamma a}'(kb)}{y_{\gamma a}(kb) + kby_{\gamma a}'(kb)} = \frac{\dfrac{\epsilon}{\epsilon_o}\dfrac{j_{\gamma a}(ka)}{j_{\gamma a}'(ka)}\left\{\dfrac{h_{\gamma a}^{(1)}(k_oa)}{h_{\gamma a}^{(1)\prime}(k_oa)} + k_oa\right\} - \dfrac{h_{\gamma a}^{(1)}(k_oa)}{h_{\gamma a}^{(1)\prime}(k_oa)}\left\{\dfrac{j_{\gamma a}(ka)}{j_{\gamma a}'(ka)} + ka\right\}}{\dfrac{\epsilon}{\epsilon_o}\dfrac{y_{\gamma a}(ka)}{y_{\gamma a}'(ka)}\left\{\dfrac{h_{\gamma a}^{(1)}(k_oa)}{h_{\gamma a}^{(1)\prime}(k_oa)} + k_oa\right\} - \dfrac{h_{\gamma a}^{(1)}(k_oa)}{h_{\gamma a}^{(1)\prime}(k_oa)}\left\{\dfrac{y_{\gamma a}(ka)}{y_{\gamma a}'(ka)} + ka\right\}}\frac{j_{\gamma a}'(ka)}{y_{\gamma a}'(ka)} \quad (25)$$

From which γ may be found in terms of the antenna geometry. For $\gamma = \beta + i\alpha$, $\beta > k_o$, $\beta \gg \alpha$ and k_oa large, a first approximation to (25) is

$$\frac{j_{\beta a}(kb) + kbj_{\beta a}'(kb)}{y_{\beta a}(kb) + kby_{\beta a}'(kb)} = \frac{\dfrac{\epsilon}{\epsilon_o}\dfrac{j_{\beta a}(ka)}{j_{\beta a}'(ka)}\left\{\dfrac{y_{\beta a}(k_oa)}{y_{\beta a}'(k_oa)} + k_oa\right\} - \dfrac{y_{\beta a}(k_oa)}{y_{\beta a}'(k_oa)}\left\{\dfrac{j_{\beta a}(ka)}{j_{\beta a}'(ka)} + ka\right\}}{\dfrac{\epsilon}{\epsilon_o}\dfrac{y_{\beta a}(ka)}{y_{\beta a}'(ka)}\left\{\dfrac{y_{\beta a}(k_oa)}{y_{\beta a}'(k_oa)} + k_oa\right\} - \dfrac{y_{\beta a}(k_oa)}{y_{\beta a}'(k_oa)}\left\{\dfrac{y_{\beta a}(ka)}{y_{\beta a}'(ka)} + ka\right\}}\frac{j_{\beta a}'(ka)}{y_{\beta a}'(ka)}. \quad (26)$$

Preliminary use of (26) should permit a reasonable computation of a root γa to (25). The procedure is to assume values for a, ϵ/ϵ_o and β/k_o which enable an evaluation of the right side of (26). A value of b is then obtained to bring the left side of (26) into balance. These values of a, b, and ϵ/ϵ_o are next used in (25) which becomes an equation in the one unknown γa. Setting $\gamma a = \beta a + \delta_1 + i\delta_2$ (as before) it can be reasoned that δ_1 and δ_2 are small and that rapid reduction of (25) is possible.

To study TE excitation of the dielectric-clad sphere, it is only necessary to interchange the roles of E and H. Referring again to Fig. 3, the fields in Region I become

$$E_\phi^{\ e} = h_{\gamma a}^{(1)}(k_or)\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}$$
$$\frac{C}{(\sin\theta)^{1/2}}e^{-i\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\}} \quad (27)$$

$$iw\mu_o H_r^e = \frac{1}{r}h_{\gamma a}^{(1)}(k_or)\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\frac{C}{(\sin\theta)^{1/2}}$$
$$\left\{i(\gamma a + \tfrac{1}{2}) + \tfrac{1}{2}\cot\theta\right\}e^{-i\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\}} \quad (28)$$

$$-iw\mu_o H_\theta^e = \left\{k_oh_{\gamma a}^{(1)\prime}(h_or) + \frac{1}{r}h_{\gamma a}^{(1)}(k_or)\right\}$$
$$\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}\frac{C}{(\sin\theta)^{1/2}}e^{-i\{wt - [(\gamma a + \frac{1}{2})\theta + \pi/4]\}} \quad (29)$$

whereas the fields in Region II are given by

$$E_\phi^{\ i} = \left\{Aj_{\gamma a}(kr) + By_{\gamma a}(kr)\right\}\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}$$
$$(\sin\theta)^{-1/2}e^{-i\{wt - [(\gamma a + \frac{1}{2})\theta +]\pi/4} \quad (30)$$

$$iw\mu_o H_r^{\ i} = \frac{1}{r}\left\{Aj_{\gamma a}(kr) + By_{\gamma a}(kr)\right\}\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}$$
$$(\sin\theta)^{-1/2}\left\{i(\gamma a + \tfrac{1}{2}) + \tfrac{1}{2}\cot\theta\right\}e^{-i\{\cdots\}} \quad (31)$$

$$-iw\mu_o H_\theta^{\ i} = \left\{k[Aj_{\gamma a}'(kr) + By_{\gamma a}'(kr)]\right.$$
$$\left. + \frac{1}{r}[Aj_{\gamma a}(kr) + By_{\gamma a}(kr)]\right\}\frac{\Gamma(\gamma a + 2)}{\Gamma(\gamma a + \frac{3}{2})}(\sin\theta)^{-1/2}e^{-i\{\ldots\}}. \quad (32)$$

Since $E_\phi^{\ i} \equiv 0$ at $r = b$,

$$B = -A\frac{j_{\gamma a}(kb)}{y_{\gamma a}(kb)}, \quad (33)$$

and thus the internal radial impedance at $r = a$ is

$$Z_r^{\ i} = \frac{-E_\phi^{\ i}}{H_\theta^{\ i}} = \frac{iw\mu_oa[j_{\gamma a}(ka) - \dfrac{j_{\gamma a}(kb)}{y_{\gamma a}(kb)}y_{\gamma a}(ka)]}{ka[j_{\gamma a}'(ka) - \dfrac{j_{\gamma a}(kb)}{y_{\gamma a}(kb)}y_{\gamma a}'(ka)] + [j_{\gamma a}(ka) - \dfrac{j_{\gamma a}(kb)}{y_{\gamma a}(kb)}y_{\gamma a}(ka)]}. \quad (34)$$

The external radial impedance at $r = a$ is

$$Z_r^e = -\frac{E_\phi^e}{H_\phi^e} = \frac{iw\mu_o a h_{\gamma a}^{(1)}(k_o a)}{k_o a h_{\gamma a}^{(1)\prime}(k_o a) + h_{\gamma a}^{(1)}(k_o a)} . \qquad (35)$$
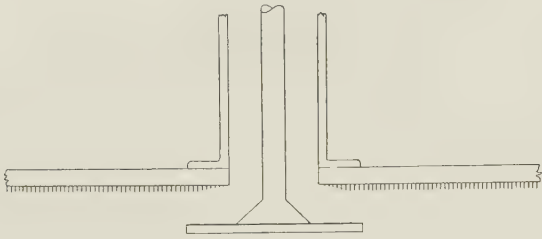
Fig. 4

The equating of (34) and (35) gives

$$\frac{j_{\gamma a}(kb)}{y_{\gamma a}(kb)} = \frac{\dfrac{k_o}{k}\dfrac{j_{\gamma a}(ka)}{j_{\gamma a}'(ka)}\dfrac{h_{\gamma a}^{(1)\prime}(k_o a)}{h_{\gamma a}^{(1)}(k_o a)} - 1}{\dfrac{k_o}{k}\dfrac{y_{\gamma a}(ka)}{y_{\gamma a}'(ka)}\dfrac{h_{\gamma a}^{(1)\prime}(k_o a)}{h_{\gamma a}^{(1)}(k_o a)} - 1}\frac{j_{\gamma a}'(ka)}{y_{\gamma a}'(ka)}, \qquad (36)$$

from which γa may be computed in terms of the physical parameters. The same computational procedure is recommended for (36) as was suggested for (25).

Fig. 6

Fig. 5

Fig. 7
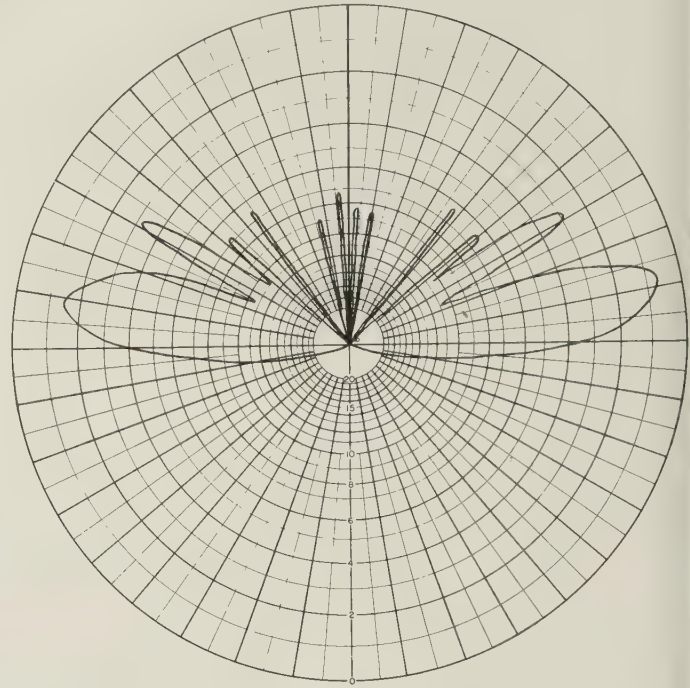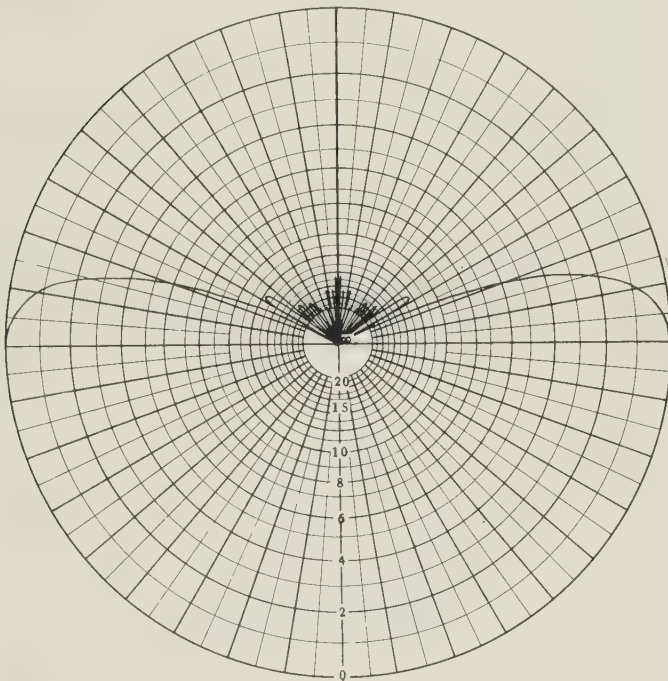
EXPERIMENT

An investigation is being conducted to study the properties of spherical cap antennas, both corrugated and dielectric-clad. To date, the results are fragmentary, but interesting.

At the beginning of the study, there was available a corrugated disk 24 inches in diameter with $G = T = 1/16$ inch and with teeth 1/16 inch deep. This was fitted with a feed as shown in Fig. 4. The feed consisted of a disk 4 inches in diameter mounted on top of a cone and excited by a coaxial line. Previous experiments had presumably determined the optimum feed dimensions. The theoretical pattern for this antenna, with an infinite ground plane, is shown in Fig. 5 and exhibits a beam-width of 17 degrees, a sidelobe level of -12 db, and a gain of 8.5 db. The experimental patterns at 9,600 mc are shown in Figs. 6 and 7. The absence of any ground
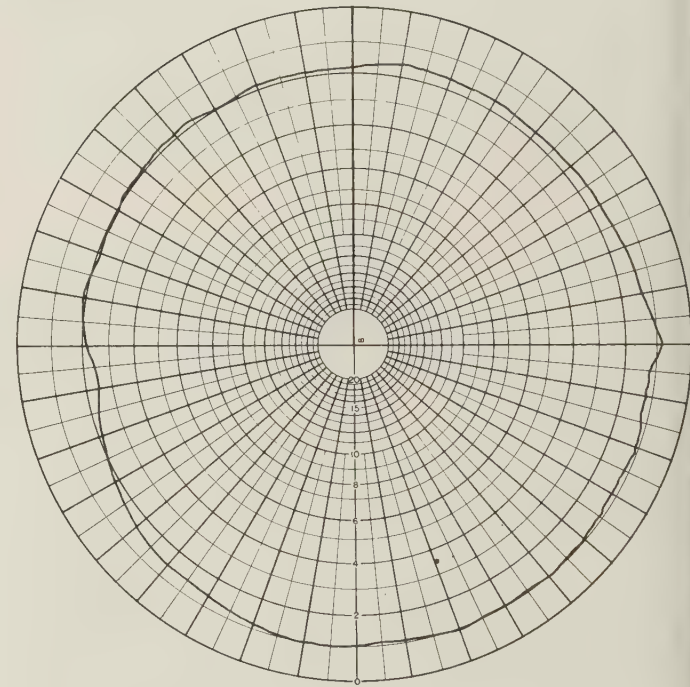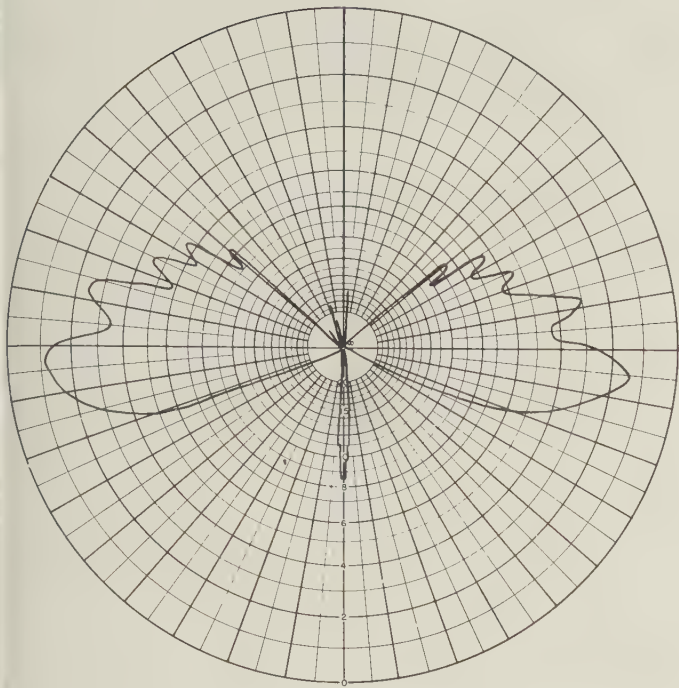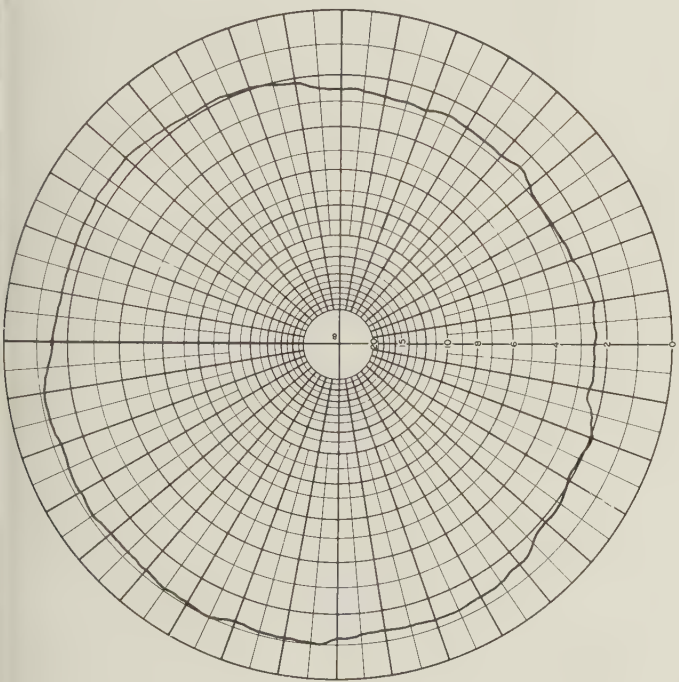
corrugated surface, the cause for the poor sidelobe figure was traced to incomplete mode conversion at the feed mouth, causing insufficiently suppressed feed radiation.

It was not the concern of this investigation to improve the performance of the corrugated disk antenna. Rather, the intention was to use it as a comparative measure of the performance of a corrugated spherical cap antenna.

Fig. 8

Fig. 9
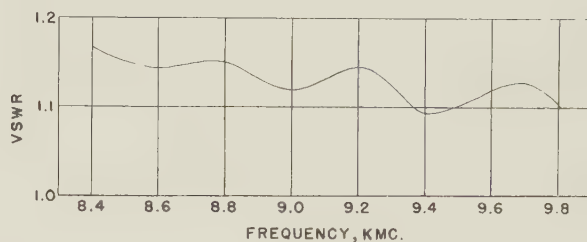
Fig. 10

To this end, a spherical surface was constructed which preserved all the dimensions of the disk except its radius of curvature, which was changed from infinity to 25.5 inches. (This radius of curvature was dictated by the fact that the maximum thickness of aluminum plate that was available was 3 inches.) The spherical cap was then fitted with the identical feed that had been used to drive the disk.

Using the theoretical procedures described earlier, the complex propagation constant of a surface wave traveling latitudinally over this spherical cap was computed to be $\gamma a = 124 + i1.4$, indicating a total leakage from the surface wave of 44 per cent of its original power content. Assuming that, to first order, the pattern from the "unleaked" part is similar to Fig. 5, and that radiation from the "leaked" part is hemispherically isotropic, the nulls of the elevation pattern could be expected to be filled in to a level of -3.5 db or better. Since the launching angle of the "unleaked" part of the wave is 28 degrees below the horizon, the beam tilt of the elevation pattern could be expected to be less than the $+11$ degrees of Fig. 6 but no less than -17 degrees.

These predictions were borne out fairly well by the experiments. Figs. 8 and 9 show the patterns for the corrugated spherical cap at 9,600 mc. Good null-filling was achieved providing solid coverage for 45 degrees of elevation above the main beam. The beam tilt was reduced 15 degrees and the conical cut still shows good circularity. The pattern gain was 5.1 db, indicating the cost of null-filling.

This antenna is extremely broadband. The pattern shape held up over a frequency range from 8,400 mc to 9,800 mc. The input vswr was measured by means of a probe inserted through an axial slit in the outer conductor of the coax, and from Fig. 10, is seen not to exceed 1.2 over the entire band.

plane causes diffraction which tilts the main beam up 11 degrees and broadens it to a 20-degree beamwidth. The sidelobe level is only -2.5 db and the gain has been reduced to 6.5 db. The conical pattern indicates that the antenna has a high degree of omnidirectionality. Except for the sidelobe level, the differences between theory and experiment are reasonable and can be explained by the absence of a ground plane. By probing the field over the

The corrugated spherical cap antenna just described is not an optimum design nor was it intended to be; its purpose has been to illustrate the theory of leaky spherical surface waves. To attain an optimum design, a more efficient feed must be devised. For a main beam placed on the horizon, the radius of curvature would have to be increased. This would restore some of the gain but would also reduce the amount of null-filling, which could in turn be restored by tapering the corrugation depths. All of this optimizing represents a considerable development effort. What the results of this first experiment appear to guarantee is that there is a sufficient dynamic range to the adjustable parameters to insure that such a development effort would be successful.

Conclusions

An analysis of latitudinal surface waves on corrugated and dielectric-clad spheres reveals the manner in which the complex propagation constant depends on radius of curvature and degree of trapping. A corrugated spherical cap antenna has been tested and offers beam-placing and null-filling advantages over its flat counterpart.

Acknowledgment

Application of Periodic Functions Approximation to Antenna Pattern Synthesis and Circuit Theory

J. C. SIMON†

Summary—Recently, mathematicians gave results on the approximation of periodic functions $f(x)$ by trigonometric sums $P_n(x)$. These results can be useful for antenna radiation and circuit theory problems. Rather than the least mean-square criterion which leads to Gibbs' phenomenon, it has been adopted that the maximum in the period of the error, $|f - P_n|$, is to be minimized. By linear transformation of the Fourier sum, a P_n sum can be obtained to give an error of the order $1/n^p$. The Fourier sum would give Log n/n^p. Limitations on the maximum of P_n derivatives are introduced allowing one to obtain the order of maximum error.

Antenna power diagram synthesis is then looked at with these results. The power radiation v^2 of an array of n isotropic independent sources equally spaced can always be written under the form of a P_n sum. Thus it is possible to give general limitations for the derivatives of v^2 in the broadside case and the endfire case. These limitations depend upon the over-all antenna dimension vs wavelength a/λ and the maximum error. A practical problem of shaped beam antenna is examined. It is shown that, by using the mathematical theory, improvements can be made on the diagram from what is usually obtained.

For circuit theory, physically evident limitations in time T and spectrum F allow one to write the most general function under the form of a P_n sum, and thus to apply the mathematical results to that field. Formal analogy allows comparison of antenna pattern and circuit theories.

1. Introduction

THE APPROXIMATION problem of functions has many applications in the field of radio-electricity. For instance, circuit synthesis and information theories use it widely. Wiener's theory of prediction and linear filtering is a famous example.[1]

In the antenna field, a major practical and theoretical problem, diagram synthesis, should also use its results.

The mathematical problem is to approximate in a certain interval a known function $f(x)$ of x with $P_n(x)$, a sum of functions of a certain type. The Fourier sums gives an example of such a procedure.

When the sum P_n is limited to a finite number of terms, an error exists between the function f and the sum P_n, except for a finite number of points.

Using a given method, the aim of the approximation theory is to make this error minimum.

Practically, in all known applications, *the total mean-square error, $\int |f - P_n|^2 dx$, is minimized.* But such a criterion is highly disputable. Usually it works very nicely with smooth, continuous functions, but not at all with steep or discontinuous ones. This comes from the fact that narrow peaks in the error function $f - P_n$ give practically no contribution to the integral of the least mean-square error. For the Fourier sums it is well-known under the name of Gibbs' phenomenon. (The Fourier sums are the sums of sine and cosine giving a minimum

mean-square error). Physicists and engineers usually consider the Gibbs' behavior as a curse impossible to escape. It is not at all exact, a sum P_n of trigonometric functions can be found without it.

Instead of the least mean-square, a much safer and simpler rule is to assume that the difference $|f - P_n|$ is inferior to a constant value *everywhere in the interval*, and to minimize this constant by choosing the proper sum P_n.

The theory of approximation by trigonometric functions, with the result that maximum for x of $|f(x) - P_n(x)|$ is minimum, has been extensively studied by mathematicians.[2-4] The next paragraphs give briefly the results and methods available.

2. The Approximation Problem

2.1 Fundamental Theorems

Theorem I (Borel): If $f(x)$ is a given periodic function of x (period equal to 2π), and $P_n(x)$ a trigonometric sum such as

$$P_n(x) = A_o + \sum_1^n A_k \cos kx + B_k \sin kx, \quad (1)$$

and if ϵ_n is the maximum for x of $|f - P_n|$ in the 2π period, then

(a) there exists a minimum of ϵ_n, say ϵ_n^o for a given n, and

(b) only one P_n gives
$$\max_x |f - P_n| = \epsilon_n^o.$$

This theorem is an "existence theorem," extremely general and good whatever the function $f(x)$ is.

Theorem II (de la Vallee Poussin):

(a) If the derivative of order p, $f^{(p)}(x)$ exists and is such that
$$|f^{(p)}(x)| < M_p,$$
then a P_n can be so found that
$$|f(x) - P_n(x)| \leqslant K_p \frac{M_p}{n^p}, \quad (2)$$

whatever the function of x and n is. K_p is an absolute constant.

(b) In reverse, if a P_n is so found that
$$|f - P_n| \leqslant \frac{B_p}{n^{p+\epsilon}} \quad (3)$$

whatever the function of x and n is, with $\epsilon > 0$ and B_p a constant, the function has a bounded derivative of order p.

† Compagnie Générale de T. S. F., Paris, France.

[1] Wiener, "Extrapolation, Interpolation...," John Wiley & Sons, Inc., New York, N. Y.; 1950.

[2] de la Vallee Poussin, "Leçons sur L'Approximation des Fonctions Continues," "Monographie Borel," Gauthiers Villars, France; 1917.
[3] Favard, "Approximation des fonctions continues," *Bull. Sc. Math. et Soc. Math. de France*; 1937.
[4] Zamansky, "Classe de saturation des procédés d'approximation," *Ann. de l'E.N.S.*; 1949–1950.

This theorem gives the order of approximation possible for a very general class of function—those which have a finite derivative. *If $|f^{(p)}| < M_p$, the order of approximation is of $1/n^p$.*

Theorem III (Bernstein): If $P_n(x)$ is a trigonometric sum of rank n [see (1)], and if max.$_x |P_n(x)| = M$, then the derivatives of $P_n(x)$ satisfy the inequalities:

$$|P_n'(x)| \leqslant nM, \quad |P_n''(x)| \leqslant n^2M, \cdots ,$$
$$|P_n^{(p)}(x)| \leqslant n^pM. \quad (4)$$

Theorem II has given the order of approximation for a very general class of functions. The order is better when a further derivative is considered. It could be thought that we have to consider a high rank of derivative to get a good approximation. But "practical functions" are very loosely defined, and Bernstein's theorem shows that for polynomials, the maximum of the derivative grows like the power of its order p. In practical applications, where functions are as "good functions" as can be, it is not of any interest to consider derivatives other than the few first ones, $f'(x)$, which represents the steepness, and $f''(x)$, which represents the skewness of the function, with perhaps the third derivative $f'''(x)$ for flatness. It can be said that these required qualities only will impose the possible approximation.

Bernstein's theorem is one of the basis of approximation theory. Besides it gives a very simple upper limit for trigonometric sum derivatives. This is often very useful to check the maximum derivative possible with a given sum of rank n.

2.2 *The Approximation Process*

Existence theorems are important, but $f(x)$ being a given function, how can be computed a P_n sum giving the predicted order of approximation n^{-p}.

Let the coefficients a_k, b_k of the Fourier series be:

$$a_k = \frac{1}{\pi} \int_o^{2\pi} f(t) \cos nt \, dt$$

$$b_k = \frac{1}{\pi} \int_o^{2\pi} f(t) \sin nt \, dt$$

$$a_o = \frac{1}{\pi} \int_o^{2\pi} f(t) \, dt. \quad (5)$$

Let $\gamma_n{}^k$ be the coefficients of a half-matrix, with the conditions $\gamma_n{}^o \equiv 1$, $\gamma_n{}^{n+k} \equiv 0$ ($k > 0$).

The sum $T_n(\gamma, x)$ is defined by

$$T_n(\gamma, x) = \frac{a_o}{2} + \sum_{k=1}^n \gamma_n{}^k(a_k \cos kx + b_k \sin kx). \quad (6)$$

The a_k and b_k are given by (5).

It has been shown (Toeplitz theorem) that under very general conditions, always realized in practice, $T_n(\gamma, x) \to f(x)$, if $n \to \infty$.

Besides mathematicians have shown that the linear γ process can always give the best order of approximation possible, predicted by general theorems.

2.3. *Remarks on Spectrum Notion and Gibb's Phenomenon*

From the physicist's point of view, the Toeplitz theorem should be of importance. It shows that the idea of spectrum has to be handled with care.

The theorem allows to form, through a linear process, an infinite number of $P_n(x)$ sums having $f(x)$ for the limit if n goes to infinity.

Allowing for the chosen error criterion, the sum $P_n(x)$ may be completely different. The Fourier sum is only the special case where the least mean-square criterion is adopted.

Thus it is possible to form the $P_n(x)$ sum, having for limit $f(x)$, even if the limit of the Fourier sum is different from $f(x)$.[5]

With functions representing a natural phenomenon, these notions are not too important. For, as will be stressed in Section 4, such functions can be exactly represented by a finite sum $P_n(x)$. The spectrum has then an unique and exact meaning. This is not true with the functions used in computations for physics. Most of these functions are of "bounded variation type." Then it is known that the limit of the Fourier series is equal to the function *except at a finite number of points.*

To this is related the "Gibbs' phenomenon." Suppose a function $f(x)$ is discontinuous at 0. In Fig. 1 $f(x)$ is in full line. There is a limit on the right A, and a limit on the left B ($OA = BO$). For $x = 0$, the limit of the Fourier series will be found equal to zero. But the limit on the right will be A', on the left B', different from A and B. This can be easily understood if the Fourier sum of n terms is represented (in dotted line on Fig. 1).
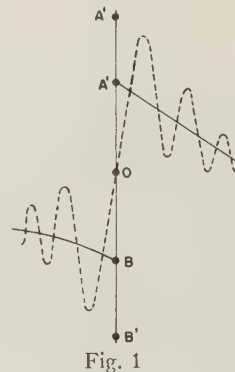
Fig. 1

The finite sum oscillates wildly around the function near the discontinuity point and goes through 0 for $x = 0$.

This is the Gibbs' phenomenon and is the consequence of the choice of the least mean-square criterion. One of the chief results of the approximation theory is to learn to find P_n sums having better behavior in such circumstances.

2.4 *Summatory Function*

If $g(t)$ is a function definite, continuous and of "bounded variation" for $0 \leqslant t \leqslant 1$, and equal to zero outside this interval, let $\gamma_n{}^k$ be defined by:

$$\gamma_n{}^k = g\left(\frac{k}{n}\right) \quad \text{with} \quad g(0) = 1, \quad (7)$$

g is said to be a summatory function.

[5] For example, see Zamansky, "La sommation des séries divergentes," "Memorial des Sciences Mathématiques," Gauthiers Villars, France, Fasc. CXXVIII; 1954.

T_n is given by:

$$T_n(g, x) = a_o + \sum_{1}^{n} g\left(\frac{k}{n}\right) [a_k \cos kx + b_k \sin kx]. \quad (8)$$

$g(t)$ being a given function, allows to build T_n in a sort of mechanical way from the Fourier series coefficients.

It has been shown that g can always be chosen so that:

$$\left| T_n(g, x) - f(x) \right| \leqslant C_p \frac{M_p}{n^p} \quad (2')$$

$$\left| f^{(p)} \right| < M_p.$$

T_n satisfies the relation (2) of the general Theorem II. *The best order of approximation is attained.*

It has been shown also that all g functions can be chosen in the family of functions defined by $g(t) = (1 - t^p)^q$. The general conclusions on the order of approximation are still good.

It is quite appealing to see the physical translation of linear process and of summatory functions. It will be seen that a "linear process" is equivalent to the operation of filtering with a linear filter having a constant transmission time or a phase-shift linear with frequency.

The spectrum deduced from the Fourier sum is transformed through a filter which has a "low-pass characteristic." The summatory function $g(k/n)$ is nothing else than the transfer function of circuit theory.

A form such as:

$$g\left(\frac{k}{n}\right) = \left[1 - \left(\frac{k}{n+1}\right)^p \right]^q$$

(with $g = 0$ for $k > n$) is quite familiar; it represents transmission through a ladder filter of q elements, each element being a maximally-flat filter with a zero of order p.

To decrease the high-order components of the spectrum has a "smoothing action" on the resulting function. Any experimenter is well aware of this fact from an intuitive point of view.

The interest of approximation theory, as it has been developed by mathematicians, is to give very general *quantitative results.* It is quite interesting to know that these mathematical theories have been worked with a scope absolutely independent of physical application. As a matter of fact French mathematicians have been completely startled that it could have any practical interest.

2.5 *The Problem of Best Constants*

Having the order of approximation, it is of interest to know the absolute constant K_p which is the best possible of all constants C_p, as used in the general Theorem II. It has been shown,[6,7] that K_p can be expressed by the formula:

$$K_p = \left(\frac{\pi}{2}\right)^p \frac{\left| E_p - D_p \right|}{p!}, \quad (9)$$

6 Favard, *Compt. Rend. Acad. Sci.*, vol. 203, p. 1122; November, 1936.

7 Norlünd, "Differenzenrechnungen," Springer, Berlin, Ger.; 1924.

with limit of K_p for $p \to \infty = 4/\pi$ and $K_p \leqslant K_1$.

E_p and D_p are Euler numbers defined by the symbolic equations:

$$(E + 2)^p + E^p = 2 \quad \text{with} \quad E_o = 1 \quad (10)$$
$$(D + 2)^p + D^p = 0 \quad \text{with} \quad D_o = 1. \quad (11)$$

It is easy to compute the first ones:

p	1	2	3
E	0	-1	1
D	-1	0	2
K	$1,57 = \dfrac{\pi}{2}$	$1,23 = \dfrac{\pi^2}{8}$	$0,64 = \dfrac{\pi^3}{48}$

Let us now examine the problem of the constant C_p of linear γ process.

If $\left| f^{(p)} \right| < M_p$, a certain class of γ process gives $T_n(\gamma, x)$, so that the error is of the order n^{-p}. The relation (2) can be written:

$$\left| T_n(\gamma, x) - f(x) \right| < \frac{M_p}{n^p} C_p(\gamma) \quad (2'')$$

$C_p(\gamma)$ being a constant which depends only on the γ process and on the order p.

Among all the γ processes, there is one, γ_o, in which $C_p(\gamma_o)$ is inferior to any other $C_p(\gamma)$. $C_p(\gamma_o)$ is called the best constant for a linear γ process. This $C_p(\gamma_o)$ is not the best of all C_p possible, such as K_p, which we examined previously. But in all practical cases, it is extremely close.

2.6 *Approximation with the Fourier Sum*

It is possible to find a P_n sum approaching a function with an error defined by formula (2). What is a Fourier sum P_n^o able to do?

If $\left| f^{(p)}(x) \right| < M_p$,

$$\left| f - P_n^o \right| < \frac{2}{\pi^2} \frac{\text{Log} n}{n^p} M_p. \quad (12)$$

The factor $2/\pi^2 \text{Log } n$ has then to compare with K_p or $C_p(\gamma)$. If n is big enough there is no question. *A P_n sum issued from a linear process will give a much lower error. It is not always the case for small enough n.*

The function $\text{Log } n$ in (12) is the mathematical expression of Gibbs' phenomenon. For small value of n it has not the possibility to build up.

2.7 *Saturation and Convergence of a γ Process*

Let, for example, $\gamma_k^n = 1 - k/n$. This is the Fejer-Cesaro sum, or first arithmetic average. It has been shown that the necessary and sufficient condition for $T(\gamma, x)$ to approach $f(x)$ with an error of the order $1/n$ is that $\left| f_1'(x) \right|$ be bounded, $f_1(x)$ being the conjugate function of $f(x)$. Even if it is possible to approximate $f(x)$ with an error the p of which is higher, as general Theorem has indicated, the Fejer process is unable to do it.

More generally a process of the form $\gamma_k{}^n = 1 - (k/n)^p$ may give $T_n - f = 0(1/n^p)$, but is unable to give better.[8] *It is said that such a process saturates at the order $1/n^p$.* The Fourier sum does not saturate. If it is possible, it gives always $0(\text{Log } n/n^p)$.

The convergence of a process is the fact that, even if the Fourier series does not converge, $T_n(\gamma, x)$ will converge towards $f(x)$. For a process with $g(t) = (1 - t^p)^q$, the convergence power increases with q. In practical cases, the Fourier series converges, and we want only to suppress the Gibbs' phenomenon. It has been shown that it is sufficient then that $g(t) = 0$ for $t = 1$. Thus the value unity will be taken for q.

2.8. *Some Remarks on Derivatives*

1. A process like $1 - t^p$, for instance, which gives $0(1/n^p)$ with a given $f(x)$, will give the same with the derivatives, if this is possible. It has to be observed that, if $f^{(r)}(x)$ is bounded, $f'(x)$ has only a derivative of order $r - 1$ bounded, and $f''(x)$ of order $r - 2$ etc. Then, $f(x)$ can be approached with $0(1/n^r)$, $f'(x)$ approached only to $0(1/n^{r-1})$, $f''(x) \cdots$ to $0(1/n^{r-2})$.

A comparison between r and p has then usually to be made to know how the different derivatives can be approached.

2. The second part of Theorem II says that if $f(x)$ is approached to $0(1/n^{p+\epsilon})$ (ϵ small) then $f^{(p)}(x)$ is bounded. This is the mathematical translation of the evident physical fact that the more a function has to be approached, the smoother it should be.

3. A limitation on $f^{(p)}(x)$ implies limitations on derivatives of $P_n(x)$ approaching $f(x)$ with $0(1/n^p)$. These sums P_n are of a more restricted class that the most general P_n sum, whose derivatives are limited by the Bernstein theorem only.

Zamansky Theorem: Let all P_n which converges uniformly towards a continuous periodic function $f(x)$[9] be considered so that:

$$\text{max.} \, | P_n(x) - f(x) | < \frac{\varphi(n)}{n^{q-1}} , \qquad (13)$$

A, B, C being absolute constants, then the derivative of rank q of P_n satisfies

$$\frac{| dP_n{}^{(q)}(x) |}{| dx^q |} < A + Bn\varphi(n) + C \int_1^n \varphi(t) \, dt, \qquad (14)$$

with q such an integer that $q \geqslant 1$, and $\varphi(t)$ being a positive, continuous, but not increasing function defined for $t \geqslant 0$.

Examples: (α) If $| P_n - f | < \epsilon$, ϵ being independent of n, then

$$| P_n' | < A + C\epsilon + (B - C)n\epsilon \qquad (15)$$

or

$$P_n' = 0(n\epsilon) \qquad (15')$$

[8] With a simplified writing, $T_n - f = 0(\psi(n))$ means that the approximation absolute error is of the order of $\psi(n)$, for example, (2) of 2.1 can be written $f - P_n = 0(1/n^p)$.

[9] To simplify the writing M the maximum of $|f(x)|$ is taken equal to 1.

$$| P_n'' | < A + \frac{C}{2} \epsilon + \left(B - \frac{C}{2} \right) n^2 \epsilon^2 \qquad (16)$$

or

$$P_n'' = 0(n^2\epsilon^2). \qquad (16')$$

(β) If $| P_n - f |$ goes to zero if n goes to infinity, then P_n'/n goes to zero if n goes to infinity, instead of $| P_n' | < n$ (*Bernstein*).

(γ) $P_n - f = 0(1/n^p)$.

All the derivatives are bounded by finite number up to the order $p - 1$, therefore:

$$P_n{}^{(p)} = 0(\text{Log } n),$$
$$P_n{}^{(p+1)} = 0(n),$$
$$P_n{}^{(p+2)} = 0(n^2) \cdots .$$

(δ) Fourier sums $P_n - f = 0(\text{Log } n/n$ $|f'(x)|$ bounded, $P_n' = 0[(\text{Log } n)^2]$, $P_n'' = 0(n \, \text{Log } n)$

2.9 *Practical Applications*

In practical applications the function $f(x)$ to be approached is usually defined in a rather loose way. The physicist is often looking for a $P_n(x)$ sum of fixed rank that meets best his desires. He can start the approximation process from many functions $f(x)$. Aware of the approximation theory, he should be able to choose function $f(x)$ to approximate, that will give him the best final results. Certain characters of $f(x)$ may have bad consequences for the approximation error, such as discontinuities or high derivatives. In practice, only the first two derivatives have to be considered. When their maximums M_1 and M_2 are known, the lowest possible errors ϵ_1, and ϵ_2 are also known. $\epsilon_1 = \pi/2(M_1/n$ $\epsilon_2 = \pi^2/8(M_2/n^2)$. It is a good rule to choose M_1 and M_2 so that $\epsilon_1 = \epsilon_2 = \epsilon$, $M_1/M_2 = \pi/4n$. Of course, the maximum error ϵ', in certain intervals smaller that 2π may be found much smaller than ϵ'. An "ideal function" $f(x)$ thus is built, and will be the basis of approximation. The first step is to compute the Fourier sum of rank n. If n is small and if $f(x)$ is a good function, the Fourier sum can be a good answer. Most of the time this is not true, and the Fourier sum exhibits the Gibbs' behavior.

A linear γ process is then used to "smooth" the Fourier sum. It has been seen that the Gibbs' behavior is suppressed if the summatory function $g(t)$ (with $t = k/n + 1$) is equal to zero for $t = 1$. Practically it is always good to choose $g(t) = 1 - t^p$ for summatory function. Such a $g(t)$ allows one to write $|f - P_n| <$ $C_p(M_p/n^p) = \epsilon_p$.

In many practical cases, the problem is to have the smallest ϵ_p, *with a given n.* But mathematics are only able to give asymptotic theorems, saying that for very great n and a given $f(x)$ the approximation will always be better with p as great as possible. For a limited n it may not be the case, because C_p and M_p increase also with p (see 2.2 Bernstein theorem). For a given n there is an optimum value of p.

It is also often of practical importance to approximate the first and sometimes the second derivative. The same principles apply then (see 2.8).

General theorems are of great help to lead the choice, but the optimum p has to be found finally in a rather experimental way for any special case.

Remarks: 1). Let $f(x)$ be considered in a region of strong variation A, B, C, D, Fig. 2. Any approximation process will give a P_n sum going through 0 the middle of BC (Jordan's theorem). It is sometimes necessary to "cover" the point C. Then instead of $f(x)$, it is
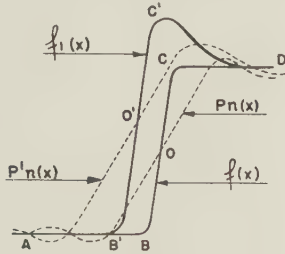

Fig. 2

better to work on $f_1(x)$, A, B', C', D. The result is then $P_n^1(x)$ going by $0'$ and which can be more satisfactory than P_n.

2. The approximation theory considers only the maximum of the error on the total interval of the 2π period. It is easy to understand that the maximum ϵ of error is likely to occur in the region of maximum variations of $f(x)$. Reciprocally, in the region where $f(x)$ is "smooth," the error can be much smaller. Such problems called "local approximation problems" are still not worked much by mathematicians. General answers are difficult to get. It is sometimes useful to approximate $\sqrt{f(x)}$, rather than $f(x)$, the error being ϵ^2 around $f(x) = 0$ and 2ϵ around $f(x) = 1$. An example is given in Section 3.

3. Diagram Synthesis with an Array of Point Sources

3.1 The problem of diagram synthesis can be limited to the two-dimensional case without great loss of generality.

If a certain wave function is assumed along a finite segment, a radiation pattern can be computed at infinity. The problem of diagram synthesis, the radiation pattern being given, is to find a wave function in the aperture. Under this condition the problem is not physically or even mathematically defined.

It is not possible to be sure that the solution is a wave function physically realizable, and the mathematical notion of function is a much too wide a conception to lead to a finite number of solutions. It is a much more physical approach to assume that a finite number n of *independent* radiation sources is available, and to see how they can be chosen to give the best possible diagram.

The given solution will at least be a solution of the wave equation, and physical realizability can easily be discussed.

3.2 *General Expression for Array Power Pattern*

Let the n sources be isotropic and equally spaced

on the x axis. In Fig. 3 the field at a remote point M is proportional to:

$$\mu = \sum_1^n v_k \cos(\varphi_k + kx - \omega t) \tag{17}$$

v_k and φ_k are the amplitude and phase of the source of order k. They completely define this source.

$x = (2\pi d/\lambda) \sin \theta$, $a = nd$, d being the spacing of sources, λ the wavelength, ω the pulsation, t the time.
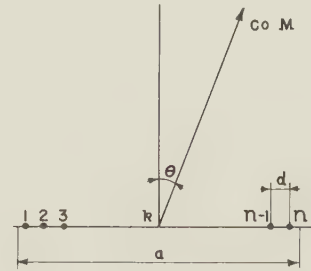
Fig. 3

Eq. (17) can be written:

$$\mu = \cos \omega t \sum_1^n v_k \cos(\varphi_k + kx)$$

$$+ \sin \omega t \sum_1^n v_k \sin (\varphi_k + kx). \tag{17'}$$

But the *only physical quantity of interest is not the instantaneous field but the average flux of power* in the vicinity of the point M. This is given by:

$$V^2 = \left[\sum_1^n v_k \cos (\varphi_k + kx) \right]^2$$

$$+ \left[\sum_1^n v_k \sin (\varphi_k + kx) \right]^2. \tag{18}$$

Expanding (18) gives:

$$V^2 = 2 \sum_{k,l=1}^n v_k v_l \cos[(\varphi_l - \varphi_k) + (l - k)x]. \tag{19}$$

Let $l - k = m$, $(k, l, m$ are integers)

$$V^2 = \frac{A_o}{2} + \sum_{m=1}^{m=n} A_m \cos mx + B_m \sin mx \tag{20}$$

with

$$\begin{cases} A_n = 2v_1 v_n \cos (\varphi_1 - \varphi_n) \\ A_{n-1} = 2v_1 v_{n-1} \cos (\varphi_1 - \varphi_{n-1}) + 2v_2 v_n \cos (\varphi_2 - \varphi_n) \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ A_m = 2 \sum_{k=1}^{k=n-m} v_k v_{k+m} \cos (\varphi_k - \varphi_{k+m}) \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ A_o = 2 \sum_1^n v^2_n \end{cases} \tag{21}$$

and

$$\begin{cases} B_n = 2v_1 v_n \sin (\varphi_1 - \varphi_n) \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ B_m = 2 \sum_{k=1}^{k=n-m} v_k v_{k+m} \sin (\varphi_k - \varphi_{k+m}) \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ B_o = 0 . \end{cases} \tag{22}$$

If the $2n$ parameters v_k, φ_k are given, the flux of energy in a direction can be written under the form of (20), the $2n + 1$ quantities A_m and B_m being given by the sets (21) and (22) in a unique way.

In reverse, if quantities A_m and B_m are given, parameters v_k and φ_k can be found by (21) and (22), but not in a unique way, as the sets are not linear. This can easily be seen by solving (21) and (22) with a decreasing order of m.

The important conclusion is *that the radiating power pattern of an array of n isotropic sources can always be represented* by a trigonometric sum P_n of rank n.

Thus the mathematical results previously exposed can be applied to that physical problem. If a certain power pattern $f^2(x)$ is given, a best power pattern V^2 can be found for a given n so that the maximum value in the interval $0, 2\pi$ of $|f^2 - V^2|$ is minimized.

Once the A_m and B_m coefficients are found, the parameters v_k, φ_k of the sources have to be determined through the sets (21) and (22). But as it has been said, the solution is not at all unique. The general equation theory shows that there are 2^n sets of solutions. Only the real roots are to be considered. Many source distributions may give the best possible power pattern. This is due to the fact that only power flux has been fixed at infinity. The phase can vary to a great extent. It would not be the case if the field itself given by its amplitude *and phase* had been predetermined. But from the physicist's and engineer's point of view, only the power is of practical interest in most cases.

3.2 Remarks

(a) The necessary and sufficient condition to have a symmetrical diagram in θ is obviously $B_m = 0$. The number of arbitrary parameters is then only $n + 1$. A sufficient condition to have all $B_m = 0$ is:

$$\begin{cases} v_k = v_{n-k} \\ \varphi_k = \varphi_{n-k} \end{cases} \quad (23)$$

The different sources are symmetrical in amplitude and phase. Another sufficient condition, more restrictive, is to have φ_k constant.

(b) The necessary and sufficient condition to have a constant phase at infinity on the sphere of center 0, center of the array, is :

$$\begin{cases} v_k = v_{n-k} \\ \varphi_k = -\varphi_{n-k} \end{cases} \quad (24)$$

This case is interesting because of the simplicity of the formulation. Let a_k and b_k be:

$$\begin{cases} a_k = v_k \sin \varphi_k \\ b_k = v_k \cos \varphi_k \end{cases} \quad (25)$$

Formula (18) can be written, in the general case:

$$V^2 = \left[\sum_1^n a_k \cos kx + b_k \sin kx \right]^2$$
$$+ \left[\sum_1^n a_k \sin kx + b_k \cos kx \right]^2. \quad (26)$$

Using (24), it is seen that one of the square brackets is null. Thus if n is even

$$\frac{V}{2} = \sum_1^{n/2} a_k \cos kx + b_k \sin kx \quad (27)$$

if n is odd

$$\frac{V}{2} = \frac{a_o}{2} + \sum_1^{\frac{n-1}{2}} a_k \cos kx + b_k \sin kx. \quad (27')$$

Now not only the power diagram, but also the field diagram can be expressed under the form of a P_n sum.

According to the simplicity of (25) compared to set (21) and (22), it is more simple to work on V than on V^2.

But the number of arbitrary parameters is only n instead of $2n$. It must be stressed that phase constancy over the sphere of center 0 is a very special condition of no physical interest. The simplicity of formulation must not hide the loss of generality.

(c) In the general case of power diagram synthesis, *i.e.* the search of the $A_k B_k$ set, it is possible that the resulting polynomial $P_n(x)$ becomes negative. This not physically possible. But this difficulty can be easily overcome.

The only regions where P_n could be negative would be the ones where f is very close to zero. For these regions, the approximation criterion is $|f - P_n| < \epsilon_n' \cdot \epsilon_n'$, smaller that ϵ_n for the 2π interval (see Section 2). Of course, if $f > \epsilon_n'$, P_n can never be negative. Then let that rule be admitted.

In the case of power diagram synthesis, the ideal function $f(x)$ to be approached should never be smaller that the approximation error ϵ_n'.

The consequences is, of course, that it is not possible to approximate zero with a better value that $2\epsilon_n'$.

3.3 The Choice of d/λ

The radiation of isotropic source arrays is undoubtedly a periodic function of $x = 2\pi d/\lambda \sin \theta$, the period being 2π. It is interesting to look at the problem with the θ coordinate. A radiation problem being always given with that coordinate.

Let us limit ourselves to *broadside radiation problems*. In all practical cases the only region of interest is $-\pi/2 < \theta < \pi/2$, for the practical sources have a forward directivity pattern. The choice of isotropic sources is, of course, a compromise for the sake of simplification, but does not change much of the conclusions due to the fact that the most interesting part of the diagram is around $\theta = 0$.

It has to be remembered that the general diagram is obtained by multiplying the diagram of point sources by the directivity of the individual source, in most cases $\cos^2 \theta$ in power.

How should d/λ be chosen? An example will show the answer.

Suppose the problem of a narrow beam has to be examined. The diagram has to approach zero outside the interval $\pm \theta_o$ (low sidelobes), and be maximum for $\theta = 0$, θ_o being small. But the approximation process is going to be done in the x coordinate. Let x_o be the

value of x for $\theta = \theta_o$, and x_1 be the value of x for $\theta = \pi/2$, $x_1 = 2\pi\, d/\lambda$).

(α) $d = \lambda/4$. The diagram in x has the same periodicity as the θ diagram. Even with point sources there is no rear radiation.

(β) $d = \lambda/2$. The period in x is the double of the θ period. $x = x_1 = \pi$ for $\theta = \pi/2$. With point sources the diagram is symmetrical forward and backward. But if directive sources are assumed, it is no more true and the real diagram has only forward radiation.

(γ) $d = \lambda$. The x period is equal to the θ period multiplied by four. $x = \pi$ for $\theta = \pi/4$. There is a peak for $\theta = \pi/2$ with point sources. But if individual directive sources are assumed, the real diagram may still be correct.

(δ) $d > \lambda$. Then many peaks may appear in the θ diagram before $\theta = \pi/2$, even with directive individual sources which, of course, are supposed to be at most of dimensions comparable to the wavelength. The function $V(x)$ has to have a peak for $x = 0$ and be as small as possible outside the interval $(-x_o, +x_o)$. But, as d/λ grows, x_o grows. And as the approximation error is related to the maximum of the derivatives, it is of interest to make d/λ as big as possible to have a good approximation.

(ϵ) $d < \lambda/4$. x_o is small, and the approximation minimum gets bigger for a decreasing d. But, unfortunately, to put *independent* radiating sources close together is physically very difficult. It is nothing else than the problem of *supergain antenna*. It has been shown to lead to a very high Q.

(η) $\lambda/2 \leqslant d \leqslant \lambda$. In most practical cases d can be chosen in the interval $(\lambda/2, \lambda)$. The fact that the $V(x)$ function has not a periodicity multiple of the $V(\theta)$ function is not a difficulty if directive individual sources with a forward directivity of $\cos^2 \theta$ are assumed. $V(\theta)$ is only of interest for $-\pi/2 < \theta < \pi/2$.

In general conclusion for a broadside pattern, d should be chosen in the interval $(\lambda/2, \lambda)$. The directivity of the sources and the degree of radiation allowed around $\theta = \pi/2$ give the value of d/λ. At any rate the $V(x)$ function has to be as small as possible outside an interval $(-x_o', +x_o)$, x_o' and x_o being smaller than π.

If $\lambda/2 < d < \lambda$, then $\dfrac{a}{\lambda} < n < \dfrac{2a}{\lambda}$.

This relation allows us to apply the approximation theory not only to n but to a/λ, the over-all dimension of the antenna, vs the wavelength.

3.4 *General Limitations of an Array of Point Sources*

Gain: A geometrical interpretation of formulas (17) and (18) can be given. The amplitude " can be considered as the projection on a fixed axis of the sum of n vectors of amplitude v_k and phase φ_k. The power V^2 is the square of the modulus A of that vector sum. A depends upon x. Besides, the average power radiated by each source is proportional to $v_k{}^2$. Thus the gain in a direction defined by $x = 2\pi\, d/\lambda \sin \theta$ is:

$$G(x) = \frac{A^2(x)}{\Sigma v_k{}^2}. \tag{28}$$

Geometrical interpretation allows us to say that the maximum value of A is obtained when all the vectors are in phase, whatever the v_k may be. Thus the maximum gain is:

$$G = \frac{(\Sigma v_k)^2}{\Sigma v_k{}^2}. \tag{29}$$

It is easy to see then, that $\Sigma v_k{}^2$ being fixed, G_o the maximum maximorum of G is obtained if all v_k are equals. $G_o = n$.

If any amplitude distribution is given, the maximum gain is $G = (\Sigma v_k)^2/\Sigma v_k{}^2$. The maximum of all is obtained when the amplitude distribution is uniform and is equal to n, the number of independent sources.

This result is obtained, of course, only with the hypothesis that the different sources are *independent* or *uncoupled*. This means that a change in one k source changes only the kth vector in the total sum.

Anyway the important result is that the maximum gain depends only upon the total number n of independent sources.

Diagram Features: Formula (20) gives the power diagram $V^2(x)$ as a $P_n(x)$ sum of rank n. It is thus possible to apply to $V^2(x)$ the Bernstein theorem (2.2) and to obtain upper limits of derivatives.

If $|P_n(x)| < M$, then the derivative of order p, $P_n{}^{(p)}$ satisfies $|P_n{}^{(p)}| < n^p M$.

But $x = 2\pi d/\lambda \sin \theta$. Knowing the maximum value M of V^2, it is possible to give an *upper limit of* $\dfrac{d^{(p)} V^2(\theta)}{d\theta^r}$

for different values of θ.

Around $\theta = 0$

$$\left| \frac{dV^2}{d\theta} \right| \leqslant \frac{2\pi a}{\lambda} M \tag{30}$$

$$\left| \frac{d^2V^2}{d\theta^2} \right| \leqslant \left(\frac{2\pi a}{\lambda} \right)^2 M \quad \text{etc} \cdots. \tag{31}$$

Around $\theta = \pi/2$

$$\left| \frac{d^2V^2}{d\theta^2} \right| \leqslant \frac{2\pi a}{\lambda} M \quad \text{etc.} \cdots. \tag{32}$$

It is interesting to see that the number of sources n does not appear in such formulas, but only $a = nd$ which is the total length of the array when it is assumed to extend $d/2$ beyond each limit source.

On the contrary of the gain limit which depends on the number n of sources and not on the antenna dimensions, the diagram limits depend only upon the over-all antenna dimension.

Formulas of the same type will be found in Section 4 for circuit theory. In both cases T or d/λ disappear in the formulation. A formal analogy can be made

between upper frequency F of the spectrum and a/λ on the one hand, and duration of the phenomenon T and λ/d on the other. *Continuous distribution has the same meaning as infinite duration.*

Formulas (31) and (32) give a limit for the skewness of diagram. If $\Delta\theta$ is the 3 db width of an array, $V^2(\theta)$ can be written:

$$V^2 \,\#\, M - \frac{(\theta - \theta_o)^2}{2} \frac{d^2V^2}{d\theta^2} \cdot$$

Then

$$(\Delta\theta)^2 \cdot \frac{d^2V^2}{d\theta^2} \,\#\, 4M$$

for a broadside type $\cdots$ $\Delta\theta > \dfrac{\lambda}{\pi a}$ (31′)

for an endfire type $\cdots$ $\Delta\theta > \sqrt{\dfrac{2\lambda}{\pi a}}.$ (32′)

An unexpected conclusion is that, for values of $a < 2\lambda$, an endfire array may have more directivity than a broadside array. It is of course no longer true for $a > 2\lambda$.

The conclusion is that there exist upper bounds, impossible to transgress, for derivatives of the power diagram.

3.5 *Limitations Related to the Approximation Error*

The results of 2.8 can be applied to the power diagram synthesis. They show that limits of V^2 derivatives, smaller than Bernstein's, have to be taken when a good approximation is wanted.

To simplify the writing, let the maximum of $V^2(\theta)$ be equal to unity. From relation (15) and (16) of Section 2, and n not too small, new limits of V^2 derivatives may be found.

broadside around $\theta = 0$ $\dfrac{dV^2}{d\theta} = 0\!\left(\dfrac{a}{\lambda}\epsilon\right),\ \dfrac{d^2V^2}{d\theta^2} = 0\!\left(\dfrac{a^2}{\lambda^2}\epsilon^2\right)$

endfire around $\theta = \dfrac{\pi}{2}$ $\dfrac{d^2V^2}{d\theta^2} = 0\!\left(\dfrac{a}{\lambda}\epsilon\right).$ (33)

As in (31), (32), n disappears. It is possible to say that *these limitations depend only upon the total dimension of the antenna vs wavelength and upon the maximum error ϵ.*

The antenna designer is well aware of such facts. He knows that to make a better fit to the ideal diagram he will have, either to ease the restrictions on the ideal diagram, or to make a bigger antenna vs wavelength.

ϵ is the maximum error for the total variation of θ. In practice it would be of interest to obtain relations such as (33), but with the "sidelobe" level ϵ'. All that can be said in general is that $\epsilon' < \epsilon$ (see 2.9, Remark 2). Work is under progress now along that line.

Relations (31), (32), and (33) do not contain n. *Thus they should stand for continuous distributions in an aperture.*

In practice, n and a/λ are of the same order, allowing for the physical fact that it would be very difficult to have independent radiating sources closer than $\lambda/2$. Thus the results of Section 2 may be applied with this restric-

tion, with n changed in a/λ in the expressions $0(\psi(n$

Of course, as it has been pointed out in 2.9, many proximation theory results are asymptotic theore specially useful for large n. They should be hand with care when one wants to work with a fixed val of n. Besides it should not be forgotten that the proximation theory error ϵ is the maximum error ov the total variation of r.

Example of Application

If, when a/λ increases, V^2 is going to converge eve where towards a given f^2, then it can be said that product $\lambda/a \cdot dV^2/d\theta$ [broadside $\theta \,\#\, 0$] decreases to z (see 2.8). Thus it is not possible to have the steepn predicted by Bernstein's theorem, when a/λ increas The two aims, convergence towards a function eve where, and steepness, act in opposite ways.

3.6 *Computation of an Array for a Shaped Beam*

The given radiation pattern is of the cosine squ type. To compare with what is usually done, the calcu tion will be made on the field and not on the pow As it has been said in 3.2, we restrict our solution to case (b) formula (24). With a symmetric amplitude, a an anti-symmetric phase, the field at infinity can expressed in the form of a P_n sum [formulas (27) a (27′)]. To work with the field rather than with the pov may be an advantage. If ϵ_1 and ϵ_2 are the errors on f around $f(x) = 0$ and $f(x) = 1$, then on $f^2(x)$ the err will be ϵ^2_1 and $2\epsilon_2$ near 0 or 1. It is a way to emphas the difference of local errors, here to lower the sidelob

The over-all dimension is equal to $a = 21\lambda$. Referri to 3.3 and to the fact that the θ pattern is rather narr the choice $d = \lambda$ is made. Thus $x = 2\pi \sin\theta$ and total number of sources is equal to twenty-one. But w the restriction made, the P_n sum to be considered only of rank $n = 10$.

Fig. 4 presents the function $f(x)$ to be approached the a curve in full line (A, C, and D). After points or D the function goes to zero with the highest possi derivatives. Two horizontal axes give the values of x a θ. Curve a_o represents the Fourier sum of rank 10, co puted on $f(x)$ directly. A strong Gibbs' behavior appea due to a high second derivative at C. Sidelobes rather high. Point A is not covered.

As has been pointed out in 2.9 $f(x)$ should be modif so that the approximation function covers that poi $f_1(x)$ represented by the b curve B, C and D is now "ideal function."

The Fourier sum is computed and a γ process applied to get a better result. The final result is giv by the curve b_4 in the dotted line. It is obviously improvement on a_o, Fig. 4.

The γ process used is defined by the summatory fu tion $g(t) = 1 - t^p$ with $t = k/11$. Fig. 5 presents t different results of these processes for the interv A, C, and D. x is plotted on the horizontal axis. T different curves are the difference in decibels betwe the desired function $f(x)$ and the different sums P_n.

a_o' is obtained from the Fourier sum computed on (c) (curve a, Fig. 4), a_2' from the sum deduced from this Fourier sum, and a process defined by the summatory function $1 - (k/n + 1)^p$ and $p = 2$. b_o' is obtained from the Fourier sum computed on $f_1(x)$ (curve b, Fig. 4), b_2', b_4', b_6' from the corresponding approximation terms with $p = 2$, 4, and 6.
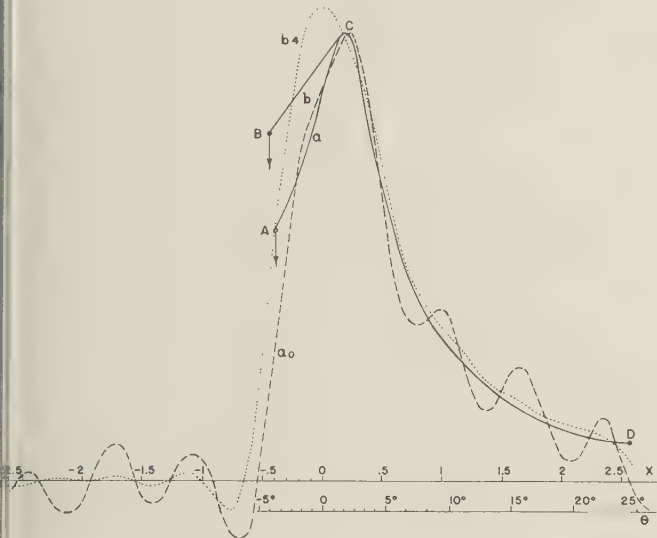

Fig. 4

Table I gives the values R in decibels of the first side-lobe level and the value x_o of x for the first zero next to point A. This last data enables us to appreciate the steepness around point A, which is of high practical importance.

TABLE I

Initial function	$f(x)$		$f_1(x)$			
p	∞	2	∞	2	4	6
R_{db}	18	25	23	34	25	24
$-x_o$	0.60	0.68	0.64	0.76	0.68	0.665

Examination of Table I and Fig. 5 shows that the most satisfactory approximation is obtained with a $1 - t^p$ process with $p = 4$.

When p increases from unity, the sidelobe level and the steepness around A increases. The limits are, of course, the sidelobe level and the steepness of the Fourier sum. On the other hand, as the approximation theory shows (see 2.9), it is of advantage to have a p as big as possible, allowing to the maximum of $f^{(p)}(x)$. For this very problem the different P_n for $p = 2$, 4, and 6 have been computed. Fig. 5 shows that the result fits best C and D, for $p = 4$, and still has low sidelobes and high steepness. It is interesting to see that $p = 2$ and $p = 6$ both give less good fit of C and D, thus allowing an experimental confirmation of the general ideas of 2.9. It has to be noted that a better fit of A and C could

have been obtained by letting the function $f_1(x)$ join more quickly $f(x)$ after it leaves the point A'. But, of course, the Fourier sum would have been more wild than the Fourier sum deduced from $f_1(x)$, and the fit on C and D less good. It is interesting also to remark that the Fourier sum of $f_1(x)$ is much quieter than the Fourier sum of $f(x)$ (see Fig. 5 curves b_o and a_o). This can be explained by the fact that the derivatives around C are then smaller.
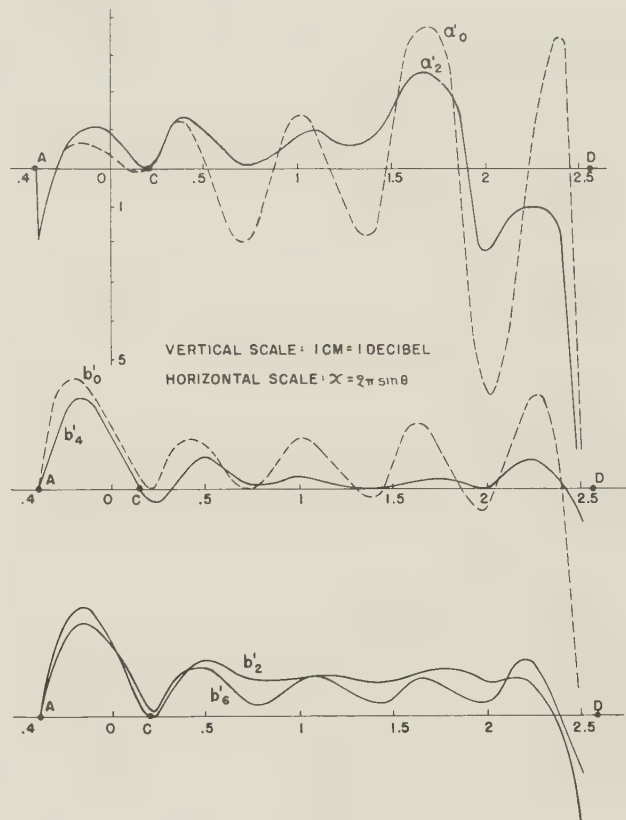
Fig. 5

The antenna physicist will not be surprised that, by filtering the Fourier sum, a smoothing action is obtained. For pencil beam patterns it is a conventional procedure. It is well-known then, that by decreasing the amplitude of the extreme sources, the sidelobe level decreases quickly without increasing much the width of the main lobe. The practical interest of the $1 - t^p$ process is to provide a tool[10] which can be used with any function $f(x)$ in a sort of mechanical way. The results of approximation can be predicted to a large extent by the mathematical theory, especially for great values of n.

4. SOME TOPICS ON CIRCUIT THEORY

Antenna field and circuit theory have many common points, mainly due to the fact that they use the same mathematical tools about periodic functions.

---

[10] As it has been said in Section 2, $1 - t^p$ gives $0(1/n^p)$ when it is possible, and good constant C_p. It is thus really not worth considering more elaborate summatory functions.

The application of the mathematical results of Section 2 gives also interesting results in the field of circuit theory. It has to be noted that the specialists of that field are already aware of the importance of approximation problems.[11]

4.1 *Introduction of Periodic Functions*

Any practical problem deals with a *finite quantity* of time T and the function is equal to zero outside this interval. It is thus always possible to consider this function as periodic and of period equal to T at most.

Let only input functions of time t periodic be considered, instead of the most general functions null from $t = -\infty$ to $t = 0$, defined for $t > 0$ up to infinity. Then the formulation of the filtering problem makes use only of trigonometric series rather than of Fourier integrals.

When it is assumed that the input function is periodic, the transient phenomenon in its generality is escaped. In fact, all difficulties due to the existence of transients disappear, such as the appearance of an output function different from zero for $t < 0$ when the input function is zero for $t < 0$.

The fact of taking care only of periodic function gives the same simplification for filter problems as the one obtained for antenna, by assuming that the aperture distribution is made of point sources instead of a continuous wave function.

4.2 *Physical Function Concept*

Another restriction, due to the physical existence possibility, very much simplifies the formulation of the problem.

Allowing for the evident fact that the physical function spectrum is limited, *any physical periodic function of circuit theory can be exactly represented by a finite trigonometric sum* such as:

$$P_n(x) = a_o/2 + \sum_1^n a_k \cos kx + b_k \sin kx. \quad (34)$$

The a_k and b_k being of course the Fourier coefficients.

This statement may seem useless. In fact, it allows us to work with finite polynomials, instead of infinite series, and to get rid of great difficulties such as the ones for the spectrum examined in 2.3.

Besides, since a physical function can always be represented exactly by a trigonometric sum, why bother to solve circuit theory problems with the most general mathematical functions? The physicist should be aware that a function in the sense of mathematics is much too wide a concept for his problems. For instance, a continuous function gives a double infinity of information in a finite interval. There is an infinity of information for one point and an infinity of points. An experimenter is only interested in a finite number of answers, *the ones*

[11] W. H. Kautz, "The approximation problem," TRANS. IRE vol. CT-1, p. 4; September, 1954.

he can check by himself. For people who like the information theory point of view, the general continuous function should be considered as a great waste. The use of a l sum allows to reduce the amount of information to finite quantity, the $2n + 1$ Fourier coefficients a_k and b.

The limitations in time domain, allowing us to use periodic function, and of frequency domain, by having finite spectrum, are quite justified for the physicist. It is thus possible to simplify greatly the statement physical problems and to be led to simple computation.

It could be thought that algebraic polynomials could be used instead of trigonometric sums. In fact, trigonometric sums are much more convenient to work with. A polynomial exists outside the finite considered interval, where it is of no interest. A consequence that, at the boundaries of the interval, they often exhibit what a physicist would call "transient behavior," which is not the case with a periodic function.

For an example of the very general conclusions obtained, let the *application of the Bernstein theorem* considered. If $P_n(x)$ is given by formula (34), and $\max._{2\pi} |P_n(x)| < M$, then $\max. |d^p P_n/dx| < n^p M$.

But a physical function of time $g(t)$ can be considered as a $P_n(x)$ polynomial. T is the duration of the phenomenon, F the maximum frequency of the Fourier spectrum when $g(t)$ is considered as periodic of period and Ω the pulsation ($\Omega = 2\pi F$). Let $x = 2\pi t/T$, $n = F \cdot T$. It is then easy to show that if

$$\max._T |g(t)| = M \quad (3$$

then the derivative of order p, $g^{(p)}(t)$ satisfies:

$$|g^{(p)}(t)| < \Omega^p M. \quad (3$$

It has to be noted that formula (36) does not mention T. It has been seen that, for a physical function $g(t)$ it is always possible to consider it as periodic with period equal to the total duration T of the phenomenon $g(t)$ being identical to zero outside. But it is often possible to find smaller intervals such that the function $g($ may be represented by a P_n sum. It is thus feasible state the following theorem:

If an interval can be found so that the function $g(t)$ can be represented in this interval by a periodic function, and M is the maximum of $|g(t)|$ in this interval, then in t interval the derivative of order p satisfies $|g^{(p)}(t)| \leqslant \Omega^p M$

This theorem gives an upper bound impossible transgress for the derivatives of a physical function.

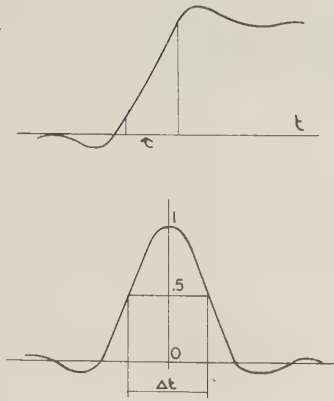(a) Rise time τ in a filter. Let τ be the rise time between 0.1 and 0.9 (Fig. 6). $1/\tau \# g'(t)$ but $|g'(t)| \leqslant \Omega$ then

$$\Omega\tau > 1. \quad (3$$

(b) Skewness of a pulse. Around the maximum (Fig. 7) $g(t) \# 1 - ((t - t_o)^2/2)g''(t_o)$ the 3 db point are defined by Δt (Fig. 7) such as $(\Delta t^2/4)g''(t_o) \#$ but $|g''(t)| < \Omega^2$, then:

$$\Omega\Delta t > 2. \quad (38$$

Relations similar to (37) and (38) are easy to find with higher derivatives. For instance, the third one concerns linearity or flatness.

All these relations are of the "Heisenberg incertitude type." The analogies with antenna problems are striking. The limitations in time and frequency for circuit problems have their equivalence in antenna radiation, *i.e.* limitations of point sources and finite length of antenna. For instance, the Bernstein theorem has the same script with $F = a/\lambda$. In both cases, n or T disappear in the final formulation.

4.3 *Linear Filter Process*

As it has been seen, any *physical* periodic function $f(x)$ can be put in a unique way under the form (34), a_k and b_k being the Fourier coefficients. This function $f(x) \equiv P_n(x)$ may represent a voltage or a current at the input of a linear filter. This linear filter excited by $f(x) = P_n(x)$ gives a voltage or current represented by a function $g(x)$ of period 2π.

A linear filter is usually defined through harmonic analysis. If the input function is a sinusoid

$$f_k{}^o(x) = a_k \cos kx + b_k \sin kx, \qquad (39)$$

the filter delivers a $g_k{}^o$ function given by

$$g_k{}^o(x) = \gamma_k a_k \cos kx + \partial_k b_k \sin kx. \qquad (40)$$

The different γ_k and ∂_k completely define the filter. In all practical cases, after a certain rank m, the coefficients are null.

$$\gamma_k \equiv \partial_k \equiv 0 \quad \text{for} \quad k > m.$$

An important practical case is the "linear phase shift" or "constant transmission time" filter. Then:

$$g_k{}^o(x) = \gamma_k f_k{}^o(x - x_o). \qquad (41)$$

x_o is independent of k and represents the time delay.

Instead of $2n + 1$ parameters, there are only n left as $\gamma_k \equiv \partial_k$.

Let $g(x)$ be the output given by a linear filter, with $f(x)$ at the input

$$g(x) = \gamma_o \frac{a_o}{2} + \sum_{k=1}^{k=m} \gamma_k a_k \cos kx + \partial_k b_k \sin kx. \qquad (42)$$

In the case of "*constant transmission time,*" $g(x)$ can be written:

$$g(x + x_o) = \gamma_o \frac{a_o}{2} + \sum_{1}^{m} \gamma_k (a_k \cos kx + b_k \sin kx). \qquad (43)$$

$g(x + x_o)$ has formally the same definition as $T_n(\gamma, x)$ of Subsection 2.2.

In that case, the filtering process is the physical expression of the "linear approximation process," as it has been defined by mathematicians.

$g(x)$ is a sum of rank equal to the smallest of the two numbers m or n. Let n be equal to m. The filter has the same rank as the input function. It is easy to see that a linear filter allows us to transform any given input function $g(x)$ into a given physical function $h(x)$ of rank p inferior or equal to n. Let $p = n$ and α_k, β_k be the $2n + 1$ Fourier coefficients, which completely define $h(x)$.

Then the $2n + 1$ equations

$$\gamma_k = \frac{\alpha_k}{a_k}, \qquad \partial_k = \frac{\beta_k}{a_k} \qquad (44)$$

completely define the solution.

Let the input function be equal to the desired function $h(x)$ plus a parasitic function $n(x)$ defined by its Fourier coefficients α_k', β_k'. Then the $2n + 1$ equations (44) become

$$\gamma_k = \frac{\alpha_k}{\alpha_k + \alpha_k'}, \qquad \partial_k = \frac{\beta_k}{\beta_k + \beta_k'}. \qquad (45)$$

The formulas (45) are very similar to the solution of filtering given by Wiener

$$Y^2(\omega) = \frac{S^2(\omega)}{S^2(\omega) + N^2(\omega)}, \qquad (46)$$

where $Y^2(\omega)$, $S^2(\omega)$, $N^2(\omega)$ are power spectrum of the filter, the signal, and the noise.

In fact for noise filtering, formulas (45) are not useful. The problem of noise filtering is to build a filter which eliminates not only one parasitic function, but a class of functions $n(x)$ statistically defined. In the case of Gaussian noise, for example, $\overline{n(x)} = 0$, and the power average $n^2(r)$ defines it completely. It is then necessary to work with functions of power, instead of voltage or current functions.

However problems of approaching a given function may be usefully investigated with the general ideas of Section 2.

CONCLUSION

In this paper two classes of results have been obtained. Theoretical ones, which are the direct application of the approximation theory of periodic functions. Practical ones, allowing the engineer to improve his computations for applications. The approach to the radiation problems by the array method appears to be the more easy to obtain general results, both on theory and

practice. Let us recall the works of Schelkunoff[12] and Dolph.[13] Approximation theories should be able to provide the physicist with theorems on radiation patterns of the same general type as the ones obtained in thermodynamics. The Bernstein theorem application is a good example. Besides, some authors, such as Bracewell[14] and Arsac,[15] have already worked on some dia-

gram synthesis for radioastronomy with different scopes than the author, independently but with similar ideas. The antenna designer should now be aware of the practical possibilities to obtain improved radiation patterns.

---

[12] Schelkunoff, "A mathematical theory of linear array," *Bell Sys. Tech. Jour.*, vol. 22, pp. 88–107; 1943.

[13] C. L. Dolph, "Current distribution for broadside . . . PROC. IRE, vol. 34, pp. 335–348; June, 1946.

[14] Bracewell, "Aerial smoothing in radioastronomy," *Aust. Jour. Phys.*, vol. 7, p. 615, 1954.

[15] Arsac, Unpublished Doctor of Science Thesis, presented in June 1955 at Paris University, France.

A Theoretical Analysis of the Multi-Element End-Fire Array With Particular Reference to the Yagi-Uda Antenna

YASUTO MUSHIAKE†

Summary—Self and mutual impedances of a multi-element antenna system are discussed, and a method of approximation for these impedances is shown. The impedances derived by this method are applied to a theoretical analysis of the multi-element parasitic end-fire array. Various characteristics of the Yagi-Uda antenna computed by the theory are given in charts, and a procedure for designing the Yagi-Uda antenna is shown. Comparisons between the theory and experimental results are also discussed.

INTRODUCTION

EVEN THE simplest antenna problem such as a single cylindrical antenna has not yet been solved exactly by the theory, whereas the multi-element end-fire array; e.g., Yagi-Uda antenna[1] which is much more complicated than the single antenna, has been working in practice for a quarter of a century.

In order to fill this large gap between the theory and practice, the author made a theoretical analysis of the multi-element end-fire array with particular reference to the Yagi-Uda antennas. Although this theory is not exact mathematically, the experimental facts for this kind of antenna can be explained by the theory, and the numerical results of this analysis can be applied for designing the Yagi-Uda antennas.

THE SELF AND MUTUAL IMPEDANCES OF AN ANTENNA ARRAY

Let us consider, at first, the simplest case of the two-element antenna array as illustrated in Fig. 1. For
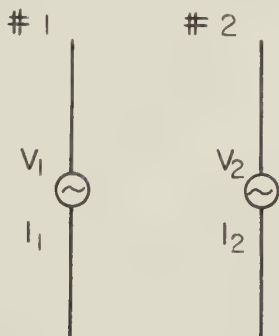

Fig. 1—Two-element array.

this system there exist well-known relations between the driving voltages V_1 and V_2 and the input currents I_1 and I_2.[2]

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \\ V_2 = Z_{21}I_1 + Z_{22}I_2 \Big\} . \qquad (1)$$

† Tohoku University, Sendai, Japan. Presently visiting Research Associate at Antenna Lab., Dept. of Elect. Engrg., Ohio State University, Columbus, Ohio.
[1] H. Yagi and S. Uda, "Projector of the sharpest beam of electric waves," *Proc. Imperial Academy of Japan*, vol. 2; 1926.
[2] P. S. Carter, "Circuit relations in radiating systems and applications to antenna problems," *Proc. IRE*, vol. 20, pp. 1004–41 June, 1932.

It must be noted here that the value of the impedance Z_{11} in this antenna system is not determined only by the dimensions of the element #1 itself as in the case of an isolated antenna. In other words, by definition, Z_{11} contains the reaction due to the induced current of element #2 when it is opened as shown in Fig. 2.
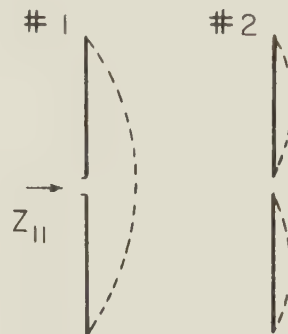
Fig. 2—Definition of self-impedance Z_{11}.

Similarly, values of the mutual impedances of a three-element antenna system contain some effects due to the existence of the third element. These facts mean that we must compute the values of these impedances considering all dimensions of the antenna system at the same time.

This causes considerable inconvenience in practical use of these exact impedances because of the large number of parameters. For this reason we made several approximations and assumptions for these values of impedances and reduced (1) to a similar one with simplified coefficients as described below.

According to the definition of the impedance in an antenna system, Z_{11} and Z_{22} can be separated into two components, respectively

$$Z_{11} = Z_{11}' + Z_{11}'' \\ Z_{22} = Z_{22}' + Z_{22}'' \Big\} , \qquad (2)$$

where Z_{11}' (or Z_{22}') corresponds to the induced voltage due to the current on element #1 (or #2), and Z_{11}'' (or Z_{22}'') corresponds to that of #2 (or #1), as shown in Fig. 2.

Now, let us consider the case where the length of the elements are nearly equal to a half wavelength and the thicknesses are very small compared with their lengths. Then it is possible to assume the following approximate relations.

$$Z_{11}' \simeq Z_{11}^{(o)} \\ Z_{22}' \simeq Z_{22}^{(o)} \Big\} , \qquad (3)$$

where $Z_{11}^{(o)}$ and $Z_{22}^{(o)}$ are the input impedance of #1 and #2, when they are isolated respectively in free

space. Furthermore, if we assume $V_2 = 0$ and not too close spacing between the elements, then we can make use of the following approximations:[3]

$$Z_{12} \simeq z_{12}m_1n_2', \qquad Z_{21} \simeq z_{21}n_1'm_2, \qquad (4)$$

$$Z_{11}''I_1 \simeq z_{12}m_1(n_2 - n_2')I_2, \qquad (5)$$

$$Z_{22}''I_2 << Z_{21}I_1, \qquad (6)$$

where $z_{12}(= z_{21})$ = the mutual impedance between two ideal half-wave dipoles which are substituted for the elements #1 and #2.

m_1, m_2 = receiving effective length factors of #1 and #2, respectively, which are defined as the ratio of the receiving voltage of the elements in free space to that of the ideal half-wave dipole.

n_1, n_2 = transmitting effective length factors of #1 and #2, respectively, which are defined as the ratio of the radiated field intensities of the elements in the equatorial plane to that of the ideal half-wave dipole with the same center current.

n_1', n_2' = special values of n_1 and n_2 for the same elements when they are driven at the center in free space and by the reciprocity theorem there are the following relations,

$$n_1' = m_1, \qquad n_2' = m_2. \qquad (7)$$

Introducing (2), (3), (4), (5) and (7) into (1) and neglecting the term $Z_{22}''I_2$ compared with $Z_{21}I_1$ in accordance with (6), we obtain

$$\left. \begin{array}{l} V_1 \simeq Z_{11}{}^{(o)}I_1 + z_{12}m_1n_2I_2 \\ 0 \simeq z_{12}m_2m_1I_1 + Z_{22}{}^{(o)}I_2 \end{array} \right\}, \qquad (8)$$

which can be transformed as follows:

$$\left. \begin{array}{l} \dfrac{V_1}{m_1} \simeq z_{11}J_1 + z_{12}J_2 \\ 0 \simeq z_{21}J_1 + z_{22}J_2 \end{array} \right\}, \qquad (9)$$

where

$$\left. \begin{array}{l} z_{11} = \dfrac{Z_{11}{}^{(o)}}{m_1n_1'} = \dfrac{Z_{11}{}^{(o)}}{m_1{}^2} \\ z_{22} = \dfrac{Z_{22}{}^{(o)}}{m_2n_2} \end{array} \right\}, \qquad (10)$$

$$\left. \begin{array}{l} J_1 = n_1'I_1 = m_1I_1 \\ J_2 = n_2I_2 \end{array} \right\}. \qquad (11)$$

Although the form of (9) is the same as that of (1), there is a large difference between these two equations. Namely, a modified self-impedance z_{11} or z_{22} is determined only by the dimensions of the element #1 or #2 itself, while the value of Z_{11} or Z_{22} depends on all the dimensions of the antenna system.

[3] S. Uda and Y. Mushiake, "Yagi-Uda Antenna," Maruzen Co., p. 65; 1954.

GENERAL EXPRESSIONS FOR THE CHARACTERISTICS OF THE N-ELEMENT PARASITIC LINEAR ARRAY

Let us consider the n-element parasitic linear array whose s-th element is placed at a distance d_s from the center of the driven element #1 as illustrated in Fig. 3. In this case the fundamental equation for the current of the elements is written as follows:

$$\sum_{s=1}^{n} Z_{rs}I_s = \begin{cases} V_1, & r = 1 \\ 0, & r \neq 1. \end{cases} \qquad (12)$$
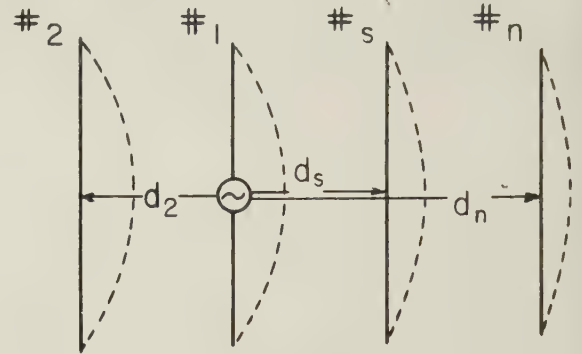

Fig. 3—n-element parasitic array.

Using the method described in the preceding section, we can reduce this equation to one which corresponds to (9); i.e.,

$$\sum_{s=1}^{n} z_{rs}J_s \simeq \begin{cases} \dfrac{V_1}{m_1}, & r = 1 \\ 0, & r \neq 1. \end{cases} \qquad (13)$$

From this equation the input impedance Z can be determined as follows:

$$Z = \dfrac{V_1}{I_1} \simeq m_1{}^2 \dfrac{[z_{rs}]}{\Delta_1} \simeq Z_{11} + \dfrac{m_1{}^2}{\Delta_1} \sum_{s=2}^{n} z_{1s}\Delta_s, \qquad (14)$$

where Δ_1, Δ_2, etc., are the minor determinants which come out in the following expansion of the determinant $[z_{rs}]$.

$$[z_{rs}] = \sum_{s=1}^{n} z_{1s}\Delta_s. \qquad (15)$$

According to the numerical computation of m_1, it is almost real, and so (14) can easily be separated into real and imaginary components and transformed to

$$R(l_1) \simeq R(\lambda/4) \mid m_1(l_1) \mid^2, \qquad (16)$$

$$X(l_1) \simeq X_{11}(l_1) + [X(\lambda/4) - X_{11}(\lambda/4)] \mid m_1(l_1) \mid^2, \qquad (17)$$

where $R(l_1)$ and $X(l_1)$ denote the values of the input resistance and the reactance of the antenna when the length of the driven element is $2l_1$.

The power gain of this antenna over that of a half-wave dipole can be calculated by the well-known procedure,[4] and it can be approximated by the following expression using (16),

$$G_{(\theta, \phi)}$$

$$\simeq \dfrac{R_o}{R(\lambda/4)} \left| \sum_{s=1}^{n} \dfrac{\Delta_s}{\Delta_1} e^{ik(d_0 \cdot ds)} \right|^2 \dfrac{\cos^2 \left(\dfrac{\pi}{2} \cos \theta \right)}{\sin^2 \theta}, \qquad (18)$$

[4] J. D. Kraus, "Antennas," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 26 and 141; 1950.

where $R_o = 73.13$ and the gain function with respect to θ is assumed to be that of a half-wave dipole.

APPLICATION TO THE MULTI-ELEMENT END-FIRE ARRAY

For the purpose of applying the above theory to the design of the multi-element antenna array it is desirable to obtain the numerical data of the characteristics of the antenna for as many combinations of the parameters as possible. Therefore, it becomes important to reduce the number of parameters in numerical computation to a minimum, without reducing the applicability of the numerical results.

The approximate expression (18) shows that the power gain of this type of antenna is almost independent of l_1. Furthermore the variation in the value of the input impedance due to the difference of l_1 from $\lambda/4$ is easily obtained from (16) and (17). Accordingly, in our numerical computation we can assume $2l_1 = \lambda/2$, and for this condition, fortunately, the effect of the thickness of #1 is almost negligible.

Next, let us consider the modified self impedance z_{22} of a parasitic element defined by (10). According to the results of the numerical computation,[5] the resistance part of the modified self impedance for a parasitic element is almost constant regardless of its length and thickness, while the imaginary part varies remarkably.
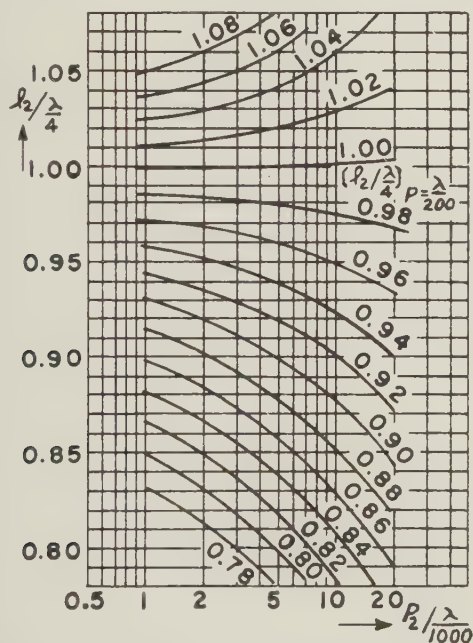
Fig. 4—Equivalent length of a parasitic element.

The curves plotted in Fig. 4 show the relation between the thickness and the length of the parasitic elements which have equal values of the modified reactance. By this chart an equivalent length of a parasitic element with any given thickness reduced to a particular reference thickness can easily be obtained. This means that without losing generality we can fix the thickness of the

parasitic elements at any convenient value in our numerical computation.

Thus, in the case of a 2-element end-fire array the number of parameters is reduced to only two; i.e., l_2 and d_2. The characteristics for this case were computed for all possible values of parameters, and the various characteristics of the director and the reflector in the Yagi-Uda antenna were explained theoretically.

By a similar procedure we computed the characteristics of a 3-element Yagi-Uda antenna and obtained many useful results. Examples of these results are shown in Fig. 5, and Figs. 6 and 7 on the next page.
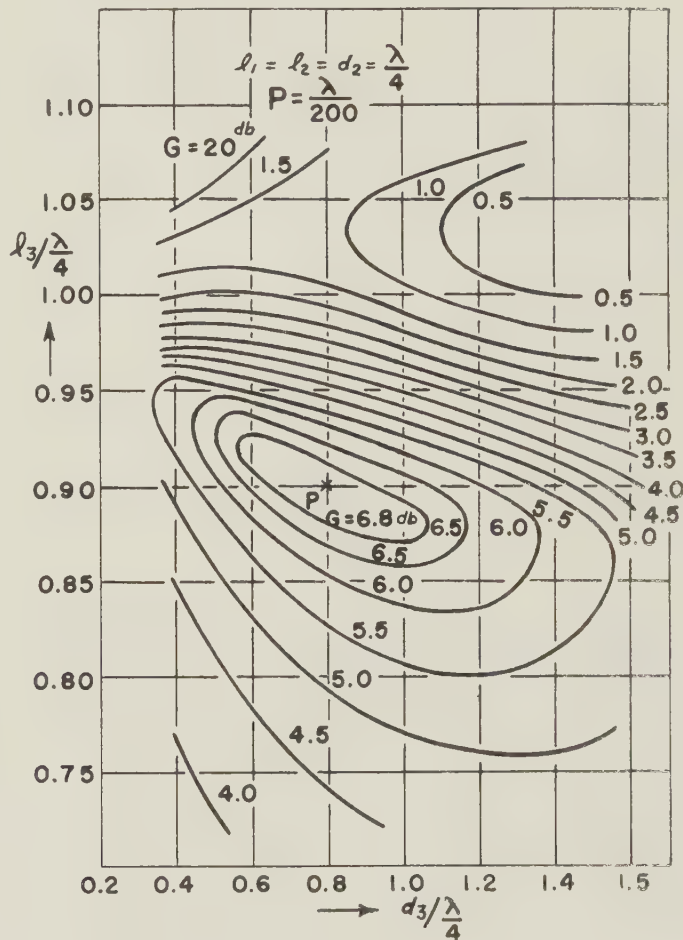
Fig. 5—Power gain of the three-element Yagi-Uda antenna.

From these results we can design a Yagi-Uda antenna with desired characteristics by the following procedure:

Design Procedure of a Yagi-Uda Antenna

1. The lengths and spacings of the directors and the reflector are determined from the chart of power gain (Fig. 5) for a reference thickness.

2. The lengths of the parasitic elements are modified for any given thickness from those of a reference thickness by Fig. 4.

3. Input reactance is cancelled by adjusting the length of driven element from the relation (17).

4. The input impedance can be increased, if necessary, to any value, by replacing the driven element with

---

[5] Uda and Mushiake, *op. cit.*, p. 75.
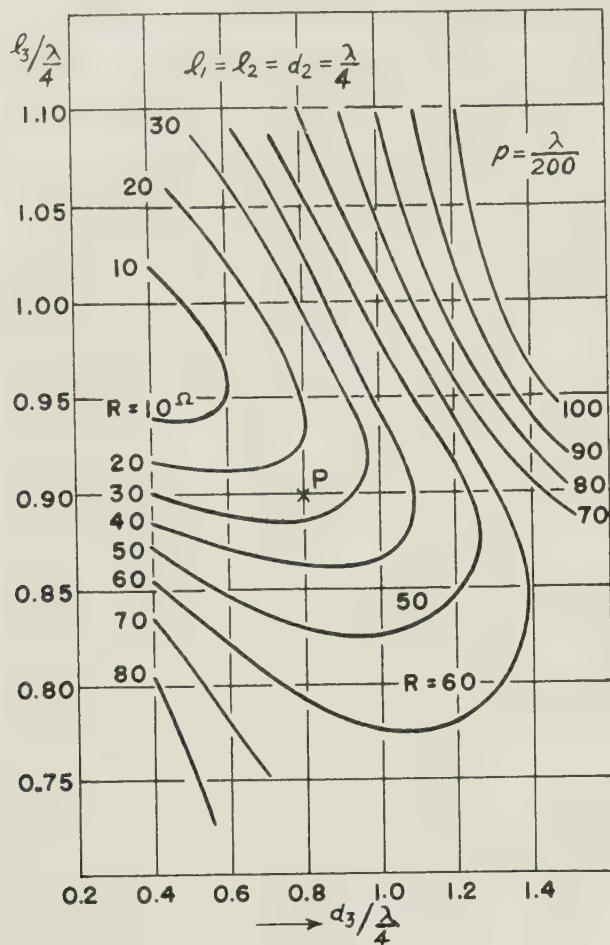
Fig. 6—Input resistance of the three-element Yagi-Uda antenna.

Fig. 7—Input reactance of the three-element Yagi-Uda antenna.

a folded dipole whose impedance transformation ratio is properly designed according to the chart given by the author.[6]

Conclusion

Since the above theory is not mathematically exact, numerous measurements[7] for more than one thousand combinations of the various parameters were made in order to verify these approximate expressions. It was found that there are good agreements between the theory and experimental results, except for a small discrepancy in the input reactances.

[6] Y. Mushiake, "An exact step-up impedance-ratio chart of a folded antenna," Trans. IRE, vol. AP-2, no. 4, p. 163; October, 1954.
[7] Uda and Mushiake, *op. cit.*, pp. 83–87, pp. 105–111, pp. 128–131 and pp. 151–160.

Although the same method of computation can be applied for any multi-element antenna array, it seems that this method is still inconvenient for obtaining a general behavior of arrays with four or more elements over the whole range of practicable values of parameters, because of the numerous combinations of parameters.

Resolution, Pattern Effects, and Range of Radio Telescopes

J. D. KRAUS†

Summary—Important source parameters and the characteristics of an ideal radio telescope are outlined. The resolution of a telescope antenna is given by Rayleigh's criterion as one-half the beamwidth between first nulls. The effect of source extent on the observed antenna pattern and the inverse problem of determining the source distribution from the measured pattern are considered. The range of a radio telescope is discussed and it is shown that some types of celestial sources could be detected far beyond the celestial horizon if such did not exist. The range of the largest optical telescopes is only half the distance to the celestial horizon, and it is pointed out that observations with large radio telescopes may be vital in determining whether a celestial horizon does in fact exist. The ultimate number of celestial sources that can be resolved with any radio telescope is given by Ko's criterion as numerically equal to the directivity of the telescope antenna.

INTRODUCTION

A RADIO telescope consists of an antenna for collecting celestial radio signals and a sensitive receiver for detecting and recording them. The antenna is analogous to the objective lens of a refracting optical telescope while the receiver-recorder is analogous to the photographic plate.[1,2]

The purpose of a radio telescope is to detect celestial radio sources. The ideal telescope should be able to provide as accurate and complete information as possible concerning each of the following source parameters: (1) position, (2) brightness distribution, (3) absolute brightness or intensity, (4) the variation of (1), (2), and (3) with time, and (5) spectrum.

The accurate *position* (right ascension and declination) of the source is necessary to distinguish the source from others, to assist in its identification with optical objects and for studies of the spatial distribution of radio sources. The accuracy of a position determination is a function not only of the resolution of the telescope but also of the source intensity, since the stronger the source the more accurately it can be located. Although no method has as yet been developed for the direct measurement of the range or distance of a radio source such information if available would be included in parameter (1).

The *brightness distribution* of a source indicates its form and extent. This information is essential for identification and classification purposes.

The *absolute value of brightness or intensity* establishes the magnitude of the source. *Brightness* (sometimes called *surface brightness*) is the power flux per unit band witdh per unit solid angle. It is expressed in janskys per square

degree (1 jansky = 1 watt per square meter per cycle per second). *Intensity* is the total power flux per unit band width as given by the integral of the brightness over the source. Thus,

$$I = \int\int B \, d\Omega, \tag{1}$$

where

I = intensity (janskys = watts meter^{-2} cps^{-1})
B = brightness (jansky deg.$^{-2}$)
$d\Omega$ = element of solid angle.

If the source extent is small compared with the beamwidth of the antenna, the intensity I is the quantity measured. To determine the absolute brightness of any part of the source requires a knowledge of the source distribution. However, if the source extent is large compared with the antenna beam width the brightness B is obtained directly from the relation

$$B = \frac{I}{\beta}, \tag{2}$$

where

B = average brightness over the beam area
I = measured intensity
β = beam area of antenna.

The *beam area* of the antenna is given by[3]

$$\beta = \int\int \frac{F(\phi, \theta)}{F_m(\phi, \theta)} \, d\Omega, \tag{3}$$

where

$F(\phi, \theta)$ = power pattern of antenna as a function of spherical co-ordinates ϕ and θ,
$F_m(\phi, \theta)$ = maximum value of $F(\phi, \theta)$
$d\Omega = \sin\theta \, d\theta \, d\phi$ = element of solid angle.

The integrand in (3) is the normalized power pattern.

It is to be noted that the directivity D of the antenna is given by[3]

$$D = \frac{4\pi}{\beta} = \frac{41253}{\beta^o}, \tag{4}$$

where

β = beam area in steradians
β^o = beam area in square degrees.

The *variation* of source parameters *with time* [item (4)] is essential to determine the constancy or variability of the radio sources. The spectrum [item (5)] concerns the variation of the other parameters as a function of the frequency. This is highly important for the classification and identification of radio sources. The window in the earth's atmosphere and ionosphere in the radio spectrum extents from at least 20 meters to 1 centimeter wavelength. Although a broad-band antenna is desirable,

[1] J. D. Kraus, "Radio telescopes," *Scientific American*, vol. 192, pp. 36–43; March, 1955.
[2] J. D. Kraus and E. Ksiazek, "New Techniques in Radio Astronomy," *Electronics*, vol. 26, pp. 148–152; September, 1953.
[3] J. D. Kraus, "Antennas," McGraw-Hill Book Co., Inc., New York, p. 24; 1950.

the 2,000 to 1 frequency range is so great as to require more than one antenna to cover it. In general, radio sources have a broad continuous spectrum with fine structure such as the hydrogen line (ground level of neutral hydrogen at 1,420 mc) superposed.

The Ideal Radio Telescope

The ideal ratio telescope is one which would measure completely all 5 parameters over a 2,000 to 1 bandwidth and also track objects across the sky from rising to setting. The telescope should also have as large a range and resolution as possible. Such a telescope is obviously not feasible and any practical design will involve compromises which will differ depending on the parameters to be measured most accurately or completely.

The signal level of celestial radio sources is very small, being of the order of 10^{-24} janskys or less, so that a large antenna aperture is desirable to increase the received power as well as improve the resolution. Although complete steerability in right ascension and declination may be desirable a much larger antenna can be constructed for the same cost with complete steerability in declination if the steerability in right ascension is restricted to angles near the meridian. Furthermore, modifications of the shape of the antenna aperture may permit the construction of apertures of much larger area for the same cost. For example, consider an antenna of square aperture (area $= L^2$) pivoted along an east-west axis located at or parallel to its lower edge, as shown in Fig. 1(a). With the lower edge at ground level and
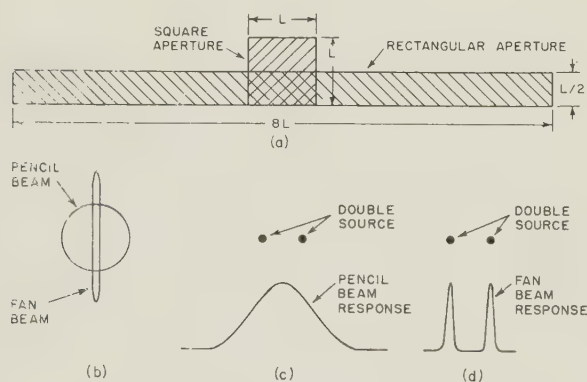

Fig. 1—Comparison of antenna with square aperture, and pencil beam and antenna with rectangular aperture and fan beam, both antennas costing the same.

assuming that the cost of such an antenna is proportional (within limits) to the cube of its height, one could obtain 4 times the aperture area for the same cost by constructing an antenna with a rectangular aperture as shown in Fig. 1(a). This antenna will have a fan beam while the square antenna beam is pencil-shaped as illustrated in Fig. 1(b). The fan beam has 8 times the resolution of the pencil beam in right ascension so that two point sources which the pencil beam is unable to resolve as shown in Fig. 1(c) can be completely resolved

by the fan beam as in Fig. 1(d). The antenna with the fan beam has twice the range of the pencil beam antenna and assuming a uniform volume distribution of sources can detect 8 times as many sources. This example illustrates how cost can be an important factor in influencing antenna design and beam shape.

Resolution

The limiting resolution of an optical device is usually given by *Rayleigh's criterion*.[4] According to this criterion two identical point sources can be just resolved if the maximum of the diffraction pattern of source 1 coincides with the first minimum of the pattern of source 2. Assuming a symmetrical antenna pattern, Rayleigh's criterion applied to antennas states that the resolution is equal to one-half the beamwidth between first nulls. That is,

$$R = \frac{\text{BWFN}}{2}, \qquad (5)$$

where

R = Rayleigh resolution or Rayleigh angle
BWFN = Beamwidth between First Nulls.

An antenna pattern for a single point source is shown in Fig. 2(a). The pattern for two identical point sources
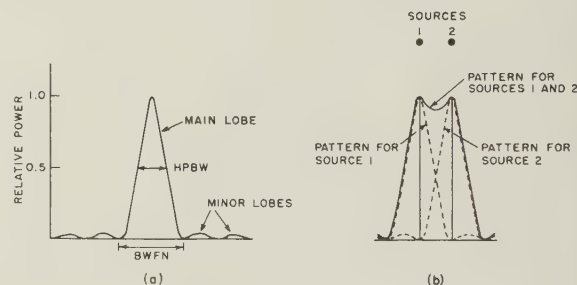
Fig. 2—(a) Antenna power pattern and (b) power patterns for two identical point sources separated by the Rayleigh angle as observed individually (dashed) and together (solid).

separated by the Rayleigh angle is given by the solid curve in Fig. 2(b), with the pattern for each source when observed individually shown by the dashed curves. It is to be noted that the two sources will be resolved provided the half-power beamwidth is less than one-half of the beamwidth between first nulls. This is usually the case.

It may be shown[3] that the beamwidth between first nulls for a broadside antenna with a uniform aperture many wavelengths long is given by

$$\text{BWFN} = \frac{114.6}{L_\lambda} \text{ degrees,} \qquad (6)$$

where

BWFN = beam width between first nulls,
L_λ = length of aperture in wavelengths.

[4] F. W. Sears, "Principles of Physics III: Optics," Addison-Wesley Press, Inc., Cambridge, Mass., p. 257; 1948.

It is assumed that $L_\lambda \gg 1$. It may also be shown that the half-power beamwidth for such an antenna is

$$\text{HPBW} = \frac{50.8}{L_\lambda} \text{ degrees.} \qquad (7)$$

These results are summarized in Table I which also gives the beamwidths for uniform circular apertures. It is to be noted that the ratio BWFN/2 HPBW is some 15 to 20 per cent greater than unity for both cases listed. Since two identical sources could just be resolved if separated by slightly more than the half-power beam width rather than the Rayleigh angle, the resolution might be taken as equal to the half-power beam width.

TABLE I

Uniform apertures	Resolution $(BWFN/2)$	Half-power beamwidth $(HPBW)$	$\dfrac{BWFN}{2\,HPBW}$
Retangular (length L_λ wavelengths)	$\dfrac{57.^\circ3}{L_\lambda}$	$\dfrac{50.^\circ3}{L_\lambda}$	1.14
Circular (diameter D_λ wavelengths)	$\dfrac{70^\circ}{D_\lambda}$	$\dfrac{59^\circ}{D_\lambda}$	1.185

However, in the interest of consistency the term *resolution* will be understood to apply to the Rayleigh angle.

A graph of the resolution or Rayleigh angle of radio and optical telescopes as a function of the telescope aperture is presented in Fig. 3.

Fig. 3—Resolution of radio and optical telescopes as a function of telescope aperture in wavelengths. Resolutions for a number of telescopes are indicated, including the pioneer radio telescopes of Karl G. Jansky and Grote Reber.

BRIGHTNESS DISTRIBUTION

The distribution of brightness across a source may, to a certain extent, be deduced from the observed antenna pattern. The *true* response pattern of a receiving antenna is obtained when the radiator is a point source situated at a su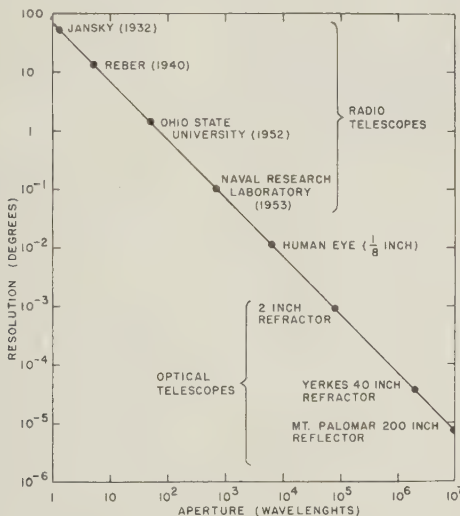fficient distance from the antenna. The distance is sufficient if an increase in the distance produces no detectable change in the pattern. Many celestial sources are of sufficient angular extent to make the observed pattern differ appreciably from the true pattern. It is one of the important problems of radio astronomy to deduce the source brightness distribution from the observed pattern.

The general problem of the effect of the source distribution on the observed antenna pattern may be stated as follows:[5]

$$G(\phi_o) = \int F(\phi + \phi_o) f(\phi) \, d\phi, \qquad (8)$$

where

$$\begin{aligned} G(\phi_o) &= \text{observed or resultant pattern} \\ F(\phi + \phi_o) &= \text{true antenna pattern} \\ f(\phi) &= \text{source (brightness) distribution} \\ \phi_o &= \text{hour angle (angle between maximum of antenna pattern and center of source).} \end{aligned}$$

The function $f(\phi)$ is equal to the brightness B in (1) and it is convenient to assume that the total intensity of the source (integral of B) is a constant.

In (8) the most general (two-dimensional) problem has been simplified to the one-dimensional case where the patterns are functions only of one co-ordinate (ϕ). This has the advantage that the basic problem is retained intact but the analysis is greatly simplified. All patterns in (8) are proportional to power.

Eq. (8) is a linear integral equation of the first kind where the true antenna pattern is the kernel and the source distribution $f(\phi)$ is the unknown function whose value is sought. Equations of this type may be solved by several methods; for example, by the method of successive substitutions. This method has been applied to the antenna problem by Bracewell and Roberts.[6] Another method of solution suggested by Calderon[7] has been applied by Matt.[5] This method involves the expansion of the patterns into orthogonal sets of functions. Thus, in the one-dimensional case of (8) the patterns are expanded into Fourier series while in the two-dimensional case the expansion is in Legendre polynomials.

A solution of practical value may also be obtained by substituting sets of plausible source distributions $f(\phi)$ into (8) and constructing families of the corresponding observed patterns $G(\phi)$. This has been done by Matt[5] for the Ohio State University radio telescope antenna by assuming uniform source distributions of various widths α. The resulting family of patterns is shown in Fig. 4 with relative received power plotted as a function of the hour angle ϕ_o.

[5] S. Matt and J. D. Kraus, "The effect of the source distribution on antenna patterns," PROC. IRE, vol. 43, pp. 821–825; July, 1955.
[6] R. N. Bracewell and J. A. Roberts, "Aerial smoothing in radio astronomy," *Australian Jour. Phys.*, vol. 7, p. 615; December, 1954.
[7] A. P. Calderon, Ohio State University, private communication; June, 1953.
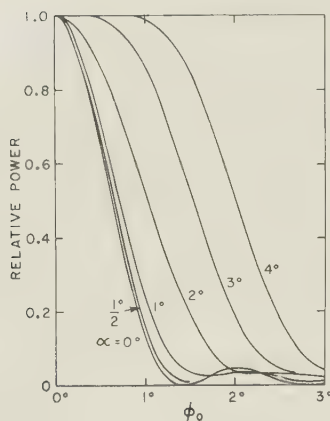
Fig. 4—Patterns which would be observed with a uniform 40-wavelength broadside antenna for uniform sources of various angular extent (α). The patterns are symmetrical, only the right half being shown.

using Fig. 6 the source width may be determined assuming it to be uniform. The actual source, of course, may not be uniform but it is a very useful procedure to specify its size in this way in terms of its *equivalent width*, that is, the width of a uniform source producing the same amount of pattern broadening at the half-power level.

The amount of precision attainable in determining the equivalent width of a source that is much narrower than the antenna resolution angle is a function of the source intensity. However, if the broadening amounts to less than a few per cent one can usually state only the maximum possible extent of the source. Thus, if the broadening observed with a 40 wavelength uniform antenna is less than one per cent it is possible to state merely that the equivalent width of the source is less than 1/10 degree.

Fig. 5—Photograph of Ohio State University radio telescope antenna. It consists of an array of 96 helical beam antennas mounted on a pivoted steel ground plane 160 feet long by 22 feet wide producing a beam approximately 1 degree wide at half power in right ascension by 8 degrees in declination at a frequency of 250 mc. The telescope receiver is housed in the building adjacent to the antenna.

A photograph of the Ohio State University radio telescope antenna is shown in Fig. 5. It is a broadside array of 96 helices arranged in 4 rows of 24 each. The rows are 40 wavelengths long at a frequency of 250 mc and the curves of Fig. 4 are for this condition. The antenna aperture is uniform so that the curves of Fig. 4 apply to any uniform broadside antenna 40 wavelengths long. The true antenna pattern is given by the curve with $\alpha = 0$ degrees. The half-power beamwidth of this pattern is 1.3 degrees and it is to be noted that for $\alpha > \frac{1}{2}$ degree there is appreciable broadening of the pattern. Fig. 6 is a graph of the amount of this broadening as a function of the source extent. By comparing an actual observed pattern with the curves of Fig. 4 or by

Fig. 6—Pattern broadening at half-power as a function of uniform source extent for a 40-wavelength uniform broadside array.

It is to be noted that even if the pattern does show appreciable broadening, details of the source distribution having a period less than a certain value will not be
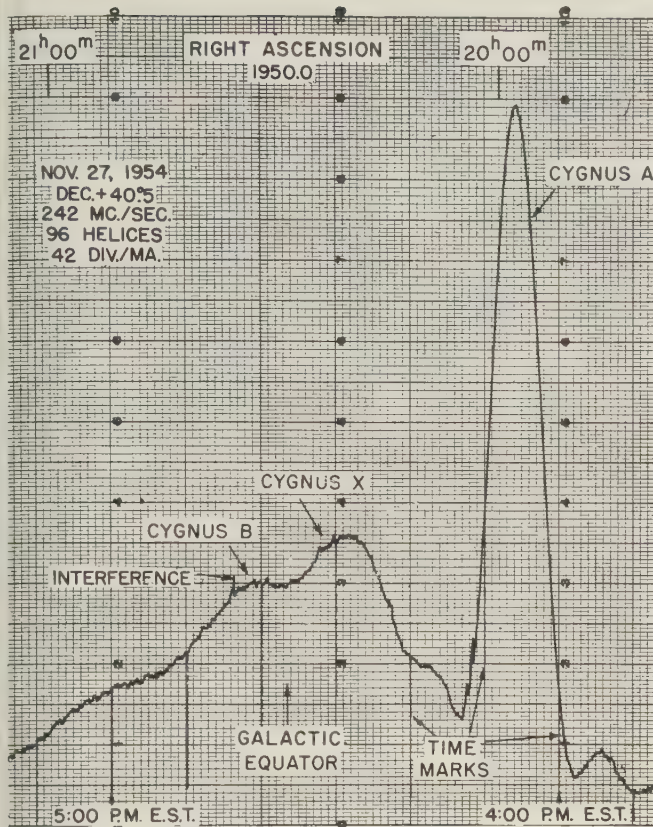
Fig. 7—Signature or profile taken with the Ohio State University radio telescope in the Cygnus region (left) and its resolution into 4 discrete sources and residual background radiation.

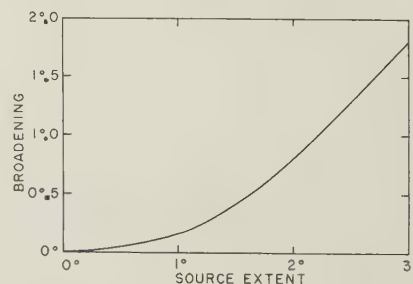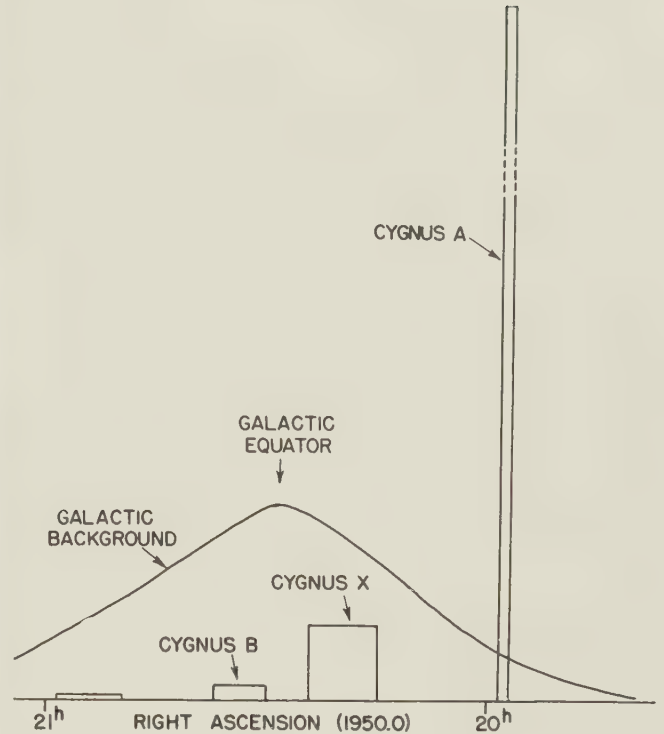observable. In this connection it may be said that the antenna produces a smoothing of the source distribution, the finer source details having no effect on the observed pattern. Conversely, it is impossible to reconstruct these details from the observed pattern in solving for the source distribution. This fact has been discussed at length by Bracewell and Roberts[6] who show that details in the source distribution having a periodicity of less than the Rayleigh angle will not appear in the observed pattern.

As an illustration of the extent to which a source distribution can be determined from an observed pattern, there is shown at the left in Fig. 7 a profile of the Cygnus region observed with the Ohio State University radio telescope at a declination of +40.5 degrees. This profile obtained by H. C. Ko on November 27, 1954 shows structure caused by a number of localized and extended radio sources. No detectable broadening is observed for the Cygnus A source but there is appreciable broadening for the sources designated Cygnus X and Cygnus B.[8] At the right in Fig. 7 is the equivalent source distribution which can be deduced from the observed profile. The Cygnus A width is known from other observations to be less than that shown but the width in Fig. 7 is an upper limit as based on the observed profile at the left. The

height of Cygnus A in the right-hand figure should be $2\frac{1}{2}$ times that shown. The weak source near right ascension 21 hours is due to the North American nebula. The removal of the 4 discrete sources from the profile leaves the smooth background level which peaks at the galactic equator. This galactic background could no doubt be resolved into a larger number of discrete sources with an antenna of higher resolution.

Telescope Range

The range of a radio telescope is of much interest since it indicates how big a sample of the universe the telescope is capable of observing. The range is a function not only of the telescope but also of the celestial sources. The celestial source having the largest known radio power output is Cygnus A which has been identified as two entire galaxies in collision at a distance of some 200 million light years.[9] It is of interest to consider the maximum distance at which this source or ones like it might be observed since this would set something of an upper limit for the telescope range.

The minimum intensity that can be detected with a radio telescope is given by

$$I_{\min} = \frac{2kTN}{A\sqrt{nt\Delta f}}, \qquad (9)$$

---

[8] See J. D. Kraus, H. C. Ko, and S. Matt," Galactic and localized source observations at 250 megacycles per second," *Astron. Jour.*, vol. 59, p. 439; December, 1954.

[9] W. Baade and R. Minkowski, "Identification of the radio sources in Cassiopeia, Cygnus A, and Puppis A," *Astrophys. Jour.*, vol. 119, p. 206; January, 1954.

where

I_{min} = minimum detectable intensity (janskys)

k = Boltzmann's constant (1.38 x 10^{-23} joules per degree Kelvin)

T = receiver ambient temperature (degrees Kelvin)

N = noise figure of receiver (dimensionless)

A = effective aperture of antenna (square meters) (includes effect of cable losses)

n = number of profiles averaged (dimensionless)

t = time constant of output integrating circuit (seconds)

Δf = radio frequency band width of receiver (cps).

The factor 2 is included since the telescope is responsive only to one polarization component of the incident radiation which is of an incoherent nature. Calculating I_{min} for the Ohio State University radio telescope operating at 250 mc, with T = 300 degrees K, N = 1.5, A = 208 square meters, t = 60 seconds, Δf = 4 mc, and for the case where 4 records are averaged (n = 4), we obtain

$$I_{min} = 1.9 \times 10^{-27} \text{ janskys.} \qquad (10)$$

The measured intensity of Cygnus A at this frequency is 38 x 10^{-24} janskys[8] so that the maximum range at which a stationary source of this type could be detected with the present Ohio State University radio telescope is given by

$$r_{max} = r_o \sqrt{\frac{I_o}{I_{min}}} = 28 \times 10^9 \text{ light years,} \qquad (11)$$

where

r_{max} = maximum range (light years)

r_o = range of Cygnus A (=200 x 10^6 light years)

I_o = intensity of Cygnus A (= 38 x 10^{-24} janskys)

I_{min} = minimum detectable intensity as in (10).

Owing to any receiver instability, the effect of which is not included in (9), and to interference and attenuation, the actual range will be less than in (11). This range, however, may be regarded as the ultimate that may be approached under ideal conditions.

The light from distant galaxies is observed to become redder the more distant the galaxy.[10] This may be explained by a Doppler shift produced by an expanding universe in which the galaxies all are receding from us, the more distant the galaxy the greater being its rate of recession. At a distance of 4 billion light years[11] the rate of recession equals the velocity of light and objects further away are beyond the range of observation. This limit to the detectable universe may be called the "celestial horizon."

The O. S. U. radio telescope has a range for stationary Cygnus A type sources of 28 billion light years (7 times the horizon distance) and hence should readily be able to detect such sources or equivalent receding ones to the celestial horizon.

This does not mean, however, that the source will be resolved since this will depend on the proximity of other sources. Actually the sensitivity of the O. S. U. radio telescope is so great that it is capable of detecting many more sources than the present antenna can resolve.

It is to be noted that the range of the largest optical telescope, the 200-inch on Mt. Palomar, is about 2 billion light years or half the distance to the celestial horizon. Radio telescopes with larger apertures than the present O. S. U. radio telescope should be able to detect and resolve sources even weaker than Cygnus-A types out to the celestial horizon and thus open to exploration vast regions of space that lie beyond the range of the largest optical telescopes. The fact that large radio telescopes should be able to detect many sources beyond the celestial horizon (if such did not exist) is of the greatest cosmological interest since a study of the intensity or magnitude distribution[12] of radio sources detected with such a telescope could be a crucial factor in determining whether a celestial horizon does in fact exist. If none were found it might mean that the universe is static and the reddening of the light from the distant galaxies is due to some other cause than a Doppler shift.

Other factors being constant the range of a radio telescope is proportional to the square root of its aperture. A graph illustrating this relation is presented in Fig. 8.
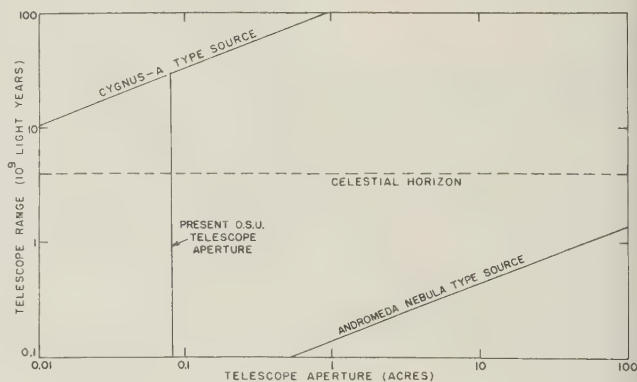
Fig. 8—Range of radio telescopes in billions of light years as a function of the area of the physical aperture of the telescope antenna in acres (1 acre = 43,560 square feet) for a Cygnus-A type source and for a normal galaxy such as the Andromeda nebula. The distance to the celestial horizon is indicated by the dashed line.

The upper slant line shows the range of radio telescopes for a Cygnus-A type source at 250 mc as based on the range of the present O. S. U. radio telescope as given in (11). The aperture in Fig. 8 is the physical aperture of the antenna which is 1.57 times the effective aperture. This is the relation for the O. S. U. radio telescope owing to about 2 db of cable loss.

The Cygnus-A radio source is two entire galaxies in collision. It is classified as an abnormal galaxy or extra-

[10] E. Hubble, "The Realm of the Nebulae," Yale University Press, 1936.

[11] All extragalactic distances in this article are on the "new scale," equal to 1.9 times the old scale. 1 light year = 6 x 10^{12} miles.

[12] J. D. Kraus, "A Preliminary Study of the Magnitude Distribution of Celestrial Radio Sources," O. S. U. Radio Observatory Rep. No. 2, Feb. 16, 1955; also Astron. Jour., in press.

[13] J. D. Kraus, "Distribution of radio brightness across the Andromeda nebula", Nature, vol. 175; p. 502; March, 1955.

alactic object. More numerous but much weaker are ne normal galaxies such as the Great Nebula in Andro- ieda (M31) which at a distance of 1.65 x 10⁶ light ears has a total intensity of 10^{-24} janskys as measured 250 mc with the O. S. U. radio telescope.[13] Introducing nese values into (11) yields 38 x 10⁶ light years as the aximum range for this type object with the O. S. U. idio telescope. To observe this type of object at the lestial horizon would require an 11,000 fold increase the antenna aperture area. Hence, until vastly larger idio telescopes are built it will not be possible to study nese weaker sources at such great distances. The lower ant line in Fig. 8 shows the range of an Andromeda ebula type of source at 250 mc as a function of antenna berture based on the calculated value for the O. S. U. idio telescope.

Ultimate Number of Sources

Assuming that a radio telescope can resolve a maxi- um of one source per beam area, the ultimate number s of celestial sources that the telescope can resolve is

$$N_s = \frac{4\pi}{\beta} = \frac{41253}{\beta_o},\qquad(12)$$

here

β = beam area of antenna (steradians)

$\beta°$ = beam area of antenna (square degrees).

he assumption of one source per beam area neglects he fact that if there is a background of extended sources ach occupying several beam areas the ultimate number buld be somewhat larger. However, (12) gives the correct

order of magnitude. Since (12) is the same relation as given by (4) for the directivity of the antenna it follows that the ultimate number of sources a radio telescope can resolve is equal numerically to the directivity D of the antenna. That is,

$$N_s = D.\qquad(13)$$

This is *Ko's criterion*, being pointed out to the author by H. C. Ko.[14] Usually N_s will be somewhat less than D since less than the entire sky is observable at most observatories.

The directivity of the Ohio State University radio telescope antenna at 250 mc is 2,850. However, objects at declinations south of -40 degrees are either below the terrestrial horizon or too close to it to permit their observation. The observable sky (north of -40 degrees) amounts to 82 per cent of the total sky so that

$$N_s = 2850 \text{ x } 0.82 = 2,340.\qquad(14)$$

This is the ultimate number of sources which can be resolved with the present O. S. U. radio telescope. This number is less than the telescope is capable of detecting. A one hundred fold increase in the area of the antenna would increase this number to 234,000.

---

[14] H. C. Ko, Ohio State University, private communication, May 1955.

Radiation from Ring Quasi-Arrays

H. L. KNUDSEN†

Summary—The present paper constitutes a summary of inves-
tigations of certain antenna systems with rotational symmetry, so-
called ring arrays and ring quasi-arrays, which have turned out to
be or can be supposed to become of practical importance.

Particular stress has been laid on an investigation of the field
radiated from homogenous ring arrays of axial dipoles and homoge-
neous ring quasi-arrays of tangential and radial dipoles; i.e.,
systems of respectively axial, tangential, and radial dipoles placed
equidistantly along a circle and carrying currents of the same
numerical value but with a phase that increases uniformly along
the circle.

At first a calculation has been made of the radiated field in the
case where the number of elements in the antenna system is
infinitely large. After that the influence of the finite number of
elements is accounted for by the introduction of correction terms.
Subsequently, the radiation resistance and the gain have been
calculated in a few simple cases.

The antenna systems described above may display super-gain.
On the basis of the theory of super-gain an estimate is made of the
smallest permissible radius of these antenna systems.

Further an investigation is made of the field from a directional
ring array with a finite number of elements to ascertain in particular
the influence on the field of the finite number of elements.

INTRODUCTION

THE PURPOSE of a transmitting antenna is to
radiate the electromagnetic energy from the
transmitter in the required manner in the various
directions of space. The radiation diagram aimed at is
often of such a kind that it is impossible to obtain the
necessary current distribution by a single antenna fed
at a single point. The desired result may then often
be obtained by using a system of separately fed an-
tennas. For practical reasons a system of identical and
identically oriented antennas, a so-called antenna array,
is often used for this purpose. The best known type of
antenna array is the linear array, i.e., an antenna array
in which the antennas are placed along a straight line,
and the linear array that has been investigated most
thoroughly in the literature is the homogeneous, linear,
array; i.e., a linear array with equidistantly placed
antennas carrying currents of the same numerical value
and with a phase that increases uniformly along the
array.

Some antennas consist of identical, but not neces-
sarily identically oriented antennas; such antenna sys-
tems will be called quasi-arrays. The present paper deals
with an investigation of the field radiated from quasi-
arrays the elements of which are placed with rotational
symmetry, so-called ring quasi-arrays; a special case
among them will be a ring array. In particular we shall
be interested in ring quasi-arrays the current distribu-
tion of which corresponds to the current distribution on
a homogeneous, linear array, so-called homogeneous
ring quasi-arrays, it being our purpose to ascertain if

such simple ring quasi-arrays have radiated fields
such a type that they can find a practical use as in t
case with the corresponding linear arrays. Further
shall investigate the field from a ring quasi-array us
as a directional beam-antenna with a principal directi
that can be altered by changing the current phases.

We are only interested in considering the field fro
ring quasi-arrays with so many elements that the fie
deviates only a little from the field found when t
number of elements is infinitely large. Therefore, in ea
case we shall start by investigating the field for an infini
number of elements. The field existing in the case of
finite number of elements will after that be express
as the field occurring when the number of elements
infinitely large plus correction terms.

The ring quasi-arrays dealt with here are suppos
to be composed of linear antennas. However, most
the characteristic features of these quasi-arrays a
found also when the linear antennas are infinitely sho
i.e., when the elements of the ring quasi-arrays a
Hertz dipoles. For this reason we start by investigati
quasi-arrays composed of Hertz dipoles and then proce
to investigate quasi-arrays of linear antennas of fini
length.

For homogeneous ring quasi-arrays the field will ha
its optimum shape when the radius of the quasi-arr
is infinitely small. Therefore we shall first investiga
the field from infinitely small ring quasi-arrays, a
next we shall see how the field is deformed when the r
dius of the quasi-array increases. Although the optimu
shape of the radiation diagram is obtained for
infinitely small value of the radius of the quasi-array,
will involve insurmountable, practical difficulties to u
a quasi-array with a very small radius. These ph
nomena will be elucidated on the basis of the theory
super-gain, and by applying this·theory we shall find t
smallest radius that it is practically possible to use
each case.

As it appears from what has been mentioned abov
it is our purpose in the present paper to consider t
fields from ring quasi-arrays having certain know
current distributions, it being our desire to investigate
quasi-arrays with current distributions of this kin
have radiated fields that make such quasi-arrays usef
in practice. Hereby we shall meet partly with anten
systems that are already known, which, however, a
here brought under a general and well-arranged calcul
tional point of view, partly with antenna systems, th
have not been utilized yet, some of which, howeve
seem to be promising from a practical point of view.

However, instead of this we could have formulat
the problem in an opposite way, namely for a give

† The Technical University of Denmark, Copenhagen, Denmark.

g quasi-array we might wish to find the current tribution that produces a desired radiated field. is problem, with which we shall not deal here, has en made the object of investigation by several authors. n calculating the radiated field from, and the radiation istance of, certain ring quasi-arrays we shall find a e for some integrals containing Bessel functions. ese integrals are expressed through known and tabu- ed functions and through the integral of a Bessel nction. A table of the integral of the Bessel function nth order has been calculated.

The present paper consists mainly of extracts from e author's treatise on antenna systems with rotational mmetry,[1] which has been published in Danish. Parts this treatise have been published in American jour- ls.[2-4] Problems that have bearing on the investigations scribed here, have been dealt with in various papers the author.[5-8]

RING ARRAY

omogeneous Ring Array of Axial Hertz Dipoles

Radiated Field; The principle of feeding the antennas a ring array with currents of the same numerical lue, but with a phase that increases uniformly along e circle; i.e. as a homogeneous ring array, seems to ve been established first by Chireix.[9] For such a ring ray with infinitely many elements Chireix has calcu- ted the radiated field and has thereby shown that this ld to an increasing degree will concentrate around the orizontal plane with increasing increment of the current ase per revolution along the circle. Apart from the articular case where the increment of the current phase zero, the ring array proposed by Chireix is therefore ited as an azimuthally omnidirectional, fading- ducing antenna. An antenna array similar to the one scribed here, but with several concentric rings, was ter proposed independently by Hansen and Wood- ard[10] and investigated further by Hansen and Hollings- orth.[11] Another type of azimuthally omnidirectional

[1] H. L. Knudsen, "Bidrag til teorien for antennesystemer med l eller delvis rotationssymmetri," I kommission hos Teknisk rlag, København; 1953.
[2] H. L. Knudsen, "The necessary number of elements in a direc- onal ring aerial," *Jour. Appl. Phys.*, vol. 22, pp. 1299–1306; 1951.
[3] H. L. Knudsen, "The field radiated by a ring quasi-array of an finite number of tangential or radial dipoles," PROC. IRE, vol. , pp. 781–789; June, 1953.
[4] H. L. Knudsen, "Radiation resistance of homogeneous ring uasi-array," PROC. IRE, vol. 42, pp. 686–695; April, 1954.
[5] H. L. Knudsen, "The field from a circular and a square helical eam antenna," *Trans. Danish Acad. Tech. Sci.*, no. 8, p. 55; 1950.
[6] H. L. Knudsen, "Radiation field of a square, helical beam ntenna," *Jour. Appl. Phys.*, vol. 23, pp. 483–491; 1952.
[7] H. L. Knudsen, "Superforstaerkning hos antenner," *Elektrotek. idsskr.*, vol. 64, pp. 213–221; 1951.
[8] H. L. Knudsen, "Shannons tidsopdelingssætning og super- orstærkning hos antenner," *Tek. Tidsskr.*, vol. 82, pp. 1023–1030; 952.
[9] H. Chireix, "Antennes a rayonnement zénithal réduit," *Onde lect.*, vol. 15, pp. 440–456; 1936.
[10] W. W. Hansen and J. R. Woodyard, "A new principle in irectional antenna design," PROC. IRE, vol. 26, pp. 333–345; larch, 1938.
[11] W. W. Hansen and L. M. Hollingsworth, "Design of "flat- nooting" antenna arrays," PROC. IRE, vol. 27, pp. 137–143; ebruary, 1939.

antenna array with reduced radiation at high elevation angles may be obtained by feeding a ring array with all currents in the same phase and adding at the center of the circle an antenna carrying a current in phase opposition to the currents in the outer antennas. This antenna array may be considered a special case of one of the arrays investigated by Hansen and Woodyard. The application of this ring antenna as a fading-reducing antenna was described by Böhm[12] and by Harbich and Hahnemann.[13]

For a homogeneous ring array the discrepancy be- tween the actual array characteristic and the array characteristic which occurs when the number of elements is infinitely large, has already been touched upon by Chireix,[9] Hansen and Woodyard,[10] and Hansen and Hollingsworth,[11] and it has later been investigated in detail by Page.[14]

Carter[15] has investigated an antenna system composed of a finite number of axial, tangential, or radial dipoles placed around a conducting cylinder, equidistantly along a concentric circle and carrying currents with the same numerical value but with a phase that increases 2π during one revolution. In the case where the radius of the conducting cylinder is zero the antenna system with axial dipoles constitutes a particular example of a ring array. In this case the expressions for the radiated field derived by Carter constitute special cases of Page's formulas.

As in the following sections we shall have use for the expression for the array characteristic of a homogeneous ring array with a finite number of elements derived by Page,[14] and as Page has not used the symbolic method for handling fields with harmonic time variation (the $-i\omega$-method) that is used throughout the following sections of this paper, we shall here give, by using this method, a short derivation of Page's results. The expres- sion for the array characteristic will here be given in a more elaborate form than in Page's work.

The homogeneous ring array considered here is sup- posed to consist of s identical and identically oriented antennas placed equidistantly along a circle with radius a. A spherical co-ordinate system (r, θ, φ) is placed with its origin at the center of the circle, which we shall here and in what follows assume to be horizontal. The co- ordinate system is supposed to be oriented in such a way that the co-ordinates of the jth antenna are given by $(a, \pi/2, u_j)$, where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots, s,$$

[12] O. Böhm, "Rundfunk-Sendeantennen mit Unterdrückung der Steilstrahlung," *Telefunkenztg*, vol. 13, pp. 21–26; March, 1932.
[13] H. Harbich and W. Hahnemann, "Wirksame Bekämpfung des Nahschwundes im Rundfunk durch Sendeantennengebilde bestimmter Form," *ENT*, vol. 9, pp. 361–376; 1932.
[14] H. Page, "Ring-aerial systems. Minimum number of radiators required," *Wireless Engr.*, vol. 25, pp. 308–315; 1948.
[15] P. S. Carter, "Antenna arrays around cylinders," PROC. IRE, vol. 31, pp. 671–693; December, 1943.

(see Fig. 1). Calling the current in the jth antenna I_j, according to the definition of a homogeneous array, we can then express I_j by

$$I_j = I e^{iHu_j},$$

where I is a constant current, and where H is an integer. The angle u_j increasing 2π when j increases from 0 to s, this equation expresses that the current phase increases linearly along the circle and just $H2\pi$ during one revolution.
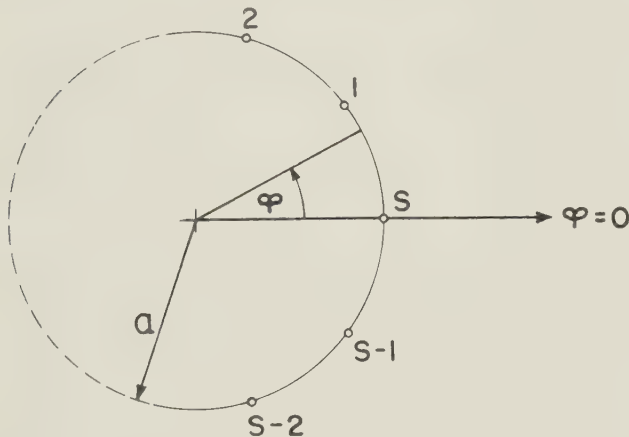

Fig. 1—Ring array with s elements and radius a.

Using as a reference antenna an antenna placed at the center of the circle, carrying the current sI; i.e., a current with the numerical value equal to the sum of the numerical values of the currents in all of the s elements of the array, we find by using the general expression for the array characteristic of an antenna array the following expression for the array characteristic of the homogeneous ring array in question,

$$G = \sum_{j=1}^{s} \frac{1}{s} e^{iHu_j} e^{-ik[a \cos u_j \sin \theta \cos \phi + a \sin u_j \sin \theta \sin \phi]}$$
$$= \frac{1}{s} \sum_{j=1}^{s} e^{i[Hu_j - z \cos (\phi - u_j)]},$$

where for the sake of convenience we have introduced the parameter z thorough the definition

$$z = ka \sin \theta.$$

The parameter z should not be mistaken for the coordinate z, which does not occur in the subsequent calculations. By using the following formula known from the theory of Bessel functions,

$$e^{iz \cos v} = \sum_{n=0}^{\infty} (2 - \delta_{on}) i^n J_n(z) \cos nv,$$

the expression for the array characteristic can be transformed into

$$G = \sum_{n=0}^{\infty} (2 - \delta_{on})(-i)^n J_n(z) \frac{1}{2s} \sum_{j=1}^{s}$$
$$[e^{i(n\phi + (H-n)u_j)} + e^{i(-n\phi + (H+n)u_j)}].$$

Both of the summations with respect to j are of the type

$$\frac{1}{s} \sum_{j=1}^{s} e^{imu_j} = \frac{1}{s} \sum_{j=1}^{s} e^{\frac{i2\pi mj}{s}} = \begin{cases} 1 \text{ for } m/s = p, \text{ where } p = \\ \qquad 0, \pm 1, \pm 2, \cdots, \\ 0 \text{ in any other case,} \end{cases}$$

where m denotes an integer. Using this formula and making the assumption

$$s > |H|,$$

we find after some rearrangement,

$$G = \frac{1}{2} \sum_{q=0}^{\infty} (2 - \delta_{oq})[J_{H+qs}(z)e^{i(H+qs)(\phi-\pi/2)}$$
$$+ J_{-H+qs}(z)e^{-i(-H+qs)(\phi+\pi/2)}$$

By multiplying this function with the field from the reference antenna, we obtain the field radiated from the antenna array in question.

Here and in what follows we shall express the electric field strength $\bar{E}(r, \theta, \varphi)$ in the field radiated from antenna system by

$$\bar{E}(r, \theta, \varphi) = K\bar{F}(\theta, \varphi) \frac{e^{ikr}}{r},$$

where $k = \omega\sqrt{\mu\epsilon}$ is the intrinsic propagation constant, K an arbitrarily chosen constant having the dimension of a voltage, and $\bar{F}(\theta, \varphi)$ a vector, being uniquely determined through the choice of K, which depends upon the direction (θ, φ), but not upon the distance r to the field point. The component of $\bar{F}$ in the direction of the r-axis is zero; consequently, denoting the unit vectors in the directions of the θ- and the φ-axes by $\hat{\theta}$ and $\hat{\varphi}$ we can express $\bar{F}$ by $\bar{F} = F_\theta\hat{\theta} + F_\varphi\hat{\varphi}$. We shall denote $\bar{F}$ as the normalized electric field strength and K as the normalization constant. The field will essentially be determined by the specification of K and $\bar{F}(\theta, \varphi)$.

The field from the reference antenna supposed to be a Hertz dipole having the length L, carrying the current sI, and placed at the center of the ring array in question is expressed by the normalization constant

$$K = \frac{iksL\zeta I}{4\pi},$$

where $\zeta = \sqrt{\mu/\epsilon}$ denotes the intrinsic impedance of space, and the normalized, electric field strength $\bar{F}$, where

$$F_\theta = -\sin \theta,$$
$$F_\varphi = 0.$$

With the normalization constant chosen here the normalized, electric field strength $\bar{F}$ of the ring array of Hertz dipoles in question will therefore be expressed by

$$F_\theta = -\frac{\sin \theta}{2} \sum_{q=0}^{\infty} (2 - \delta_{oq})[J_{H+qs}(z)e^{i(H+qs)(\phi-\pi/2)}$$
$$+ J_{-H+qs}(z)e^{-i(-H+qs)(\phi+\pi/2)}], \qquad F_\varphi =$$

We shall start by considering the field in the case where the number of elements is infinitely large. Accordingly we let $s \to \infty$ at the same time, letting the length L of the single dipoles converge towards zero in such a way that sL converges towards a finite value. Hereby we find the following limiting values of the components of the normalized, electric field strength

$$F_\theta = -\sin \theta J_H(z)e^{iH(\phi-\pi/2)},$$
$$F_\varphi = 0.$$

This special result was derived already by Chireix. It follows from the expression obtained for $\bar{F}$ that the

numerical value of the field does not depend on φ. For $|H| \geqq 1$ points in the horizontal plane having the same phase will be situated on an Archimedes' spiral as is shown in Fig. 2 for respectively $H = 1$ (full line) and $H = -1$ (dotted line). Sandeman[16] and Granquist[17] have suggested utilizing this fact for navigational purposes.
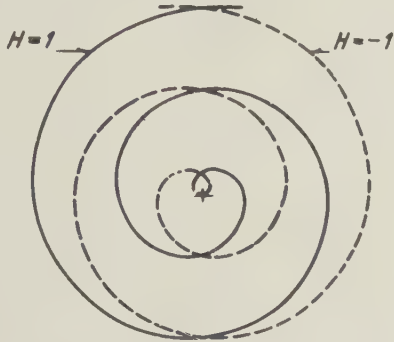

Fig. 2—Equi-phase lines for a homogeneous ring quasi-array with $H = \pm 1$.

If the radius of the ring array is much smaller than the wavelength λ divided by 2π; i.e., if $ka \ll 1$, the Bessel function in the expression given above can be expressed with sufficient accuracy by the first term of its series

$$F_\theta \cong -\frac{(ka)^H}{2^H H!} \sin^{H+1}\theta \ e^{iH(\varphi - \pi/2)},$$

$$F_\varphi = 0.$$

In Fig. 3(a)–3(d) the numerical value of F_θ is plotted as a function of θ for various values of H. The figure shows that the field, which everywhere is polarized in the θ-direction, to an increasing degree will concentrate around the horizontal plane when the increment of the current phase per revolution, as expressed by H, increases. As pointed out by Chireix, for $|H| \geqq 1$ the ring array described here is therefore suited as a fading-reducing antenna for broadcasting purposes. The homogeneous ring array with $H = 0$; i.e., with all the currents in phase, is not suited as a fading-reducing antenna, this array having the same polar diagram as a single dipole. However, as was mentioned above, when combining it with a center antenna, it is possible to obtain an antenna array having zero radiation at a given elevation angle and consequently reduced radiation at high elevation angles. Large antenna systems of this type are at present being planned by the Swedish Telegraph Administration and by Nordwestdeutscher Rundfunk in Western Germany.

If the radius of the ring array is not infinitely small, as was assumed above, the field will not assume the ideal shape shown in Fig. 3, but deviate from it to an increasing degree as the radius increases. However, if we choose as small a radius as the consideration of avoiding super-gain permits, the polar diagrams will deviate only to a small degree from those shown in Fig. 3.

So far we have assumed that the number of elements is infinitely large. However, in practice the number of elements will be finite; this implies that the antenna system will no longer be exactly azimuthally omnidirectional as in the case of infinitely many elements. Economical considerations usually make it desirable to reduce the number of elements as much as possible. It is therefore of interest to investigate the influence of the number of elements on the shape of the radiated field. In order that the field from the homogeneous ring array in question with a finite number of elements may be a fair approximation to the field from the corresponding array with infinitely many elements, the term corresponding to $q = 0$, which, as shown above, is identical with the field in the case of $s = \infty$, must be the dominant term in the infinite series. The approximate condition that this will occur is that

$$s > 2|H|;$$
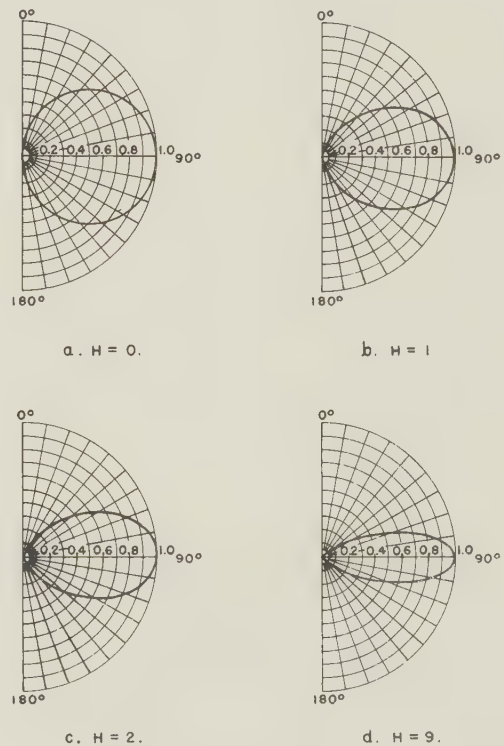
Fig. 3—The field radiated from a small, homogeneous ring array of axial dipoles for various values of H.

the Bessel function occurring in the term corresponding to $q = 0$, then, is of a smaller order than any of the Bessel functions appearing in the correction terms. In what follows we shall assume that this inequality is satisfied, and not only the less strong inequality introduced above, $s > |H|$. Subsequently we shall consider only such cases where the correction terms are small as compared with the principal term; apart from the main term it will then be necessary in general to include only the first correction term. The field is then approximately expressed by

$$F_\theta \cong -\sin\theta \ e^{iH\phi}\{(-i)^H J_H(z) + (-i)^s[(-i)^H e^{is\phi} J_{H+s}(z) + (-i)^{-H} e^{-is\phi} J_{-H+s}(z)]\}.$$

[16] E. K. Sandeman, "Spiral-phase fields," *Wireless Engr.*, vol. 6, pp. 96–105; 1949.

[17] C. E. Granquist, "Radiofyr för avstånd och riktning," Paper presented at the Radiovetenskapliga Konferensen, Stockholm; 1949.

Whereas the numerical value of the array characteristic in the case of $s = \infty$ is independent of φ, as was mentioned above, the characteristic for the array considered here will oscillate weakly with φ for any fixed value of θ.

For a function $f(\varphi)$ that oscillates weakly between two boundaries $f_{\min}$ and $f_{\max}$; i.e.,

$$f_{\min} \leqq f(\varphi) \leqq f_{\max},$$

where

$$2\delta f = f_{\max} - f_{\min} \ll f(\varphi),$$

we shall introduce for use here and in the following sections a parameter, the so-called relative, maximum variation Δf, to describe the relative change of the function

$$\Delta f = \frac{f_{\max} - f_{\min}}{(f_{\max} + f_{\min})} \underset{\sim}{=} \frac{f_{\max} - f_{\min}}{2f_{\max}} \simeq \frac{f_{\max} - f_{\min}}{2f_{\min}}.$$

In calculating the relative, maximum variation $\Delta| F_\theta |$ of the field radiated from the ring array in question we must consider the cases $H = 0$ and $| H | \geqq 1$ separately.

$H = 0$. The approximate expression for the field derived above may in the present case be written as follows

$$F_\theta \cong -\sin\theta[J_o(z) + (-i)^s 2J_s(z)\cos s\varphi].$$

Here we must treat the cases of s even and s odd separately.

s even. In the case of s even the field may be expressed with advantage by

$$F_\theta \cong -\sin\theta[J_o(z) + (-1)^{s/2} 2J_s(z)\cos s\varphi].$$

This expression shows that the correction term is in phase with or in phase opposition to the principal term for any value of φ. From this we find for the relative, maximum variation $\Delta| F_\theta |$,

$$\Delta| F_\theta | \cong \frac{2| J_s(z) |}{| J_o(z) |}.$$

s odd. When s is odd, it is convenient to transform the expression for the field as follows,

$$F_\theta = -\sin\theta[J_o(z) + i(-1)^{s+1/2} 2J_s(z)\cos s\varphi].$$

This expression shows that in the present case the correction term is in phase quadrature with the principal term for any value of φ. Therefore in the present case the relative, maximum variation $\Delta| F_\theta |$, is expressed by

$$\Delta| F_\theta | \simeq \frac{(J_s(z))^2}{(J_o(z))^2}.$$

By comparing the expressions for $\Delta| F_\theta |$ derived here for the cases of s even and s odd we see that for the same ratio of the numerical value of the correction term to the numerical value of the principal term, the relative, maximum variation of the field is essentially smaller in the case of s odd than in the case of s even. This fact appears also from Fig. 4, where the relative, maximum variation for a homogeneous ring array with $H = 0$; i.e., with currents in the same phase, is plotted as a function of $a/\lambda \sin\theta$. The curves show that three antennas produce a less wavy field than four antennas,

five antennas a better field than six, and seven antennas a better field than eight. This peculiar and practically important fact, which has not always been taken into account, was pointed out by Page.[14]

$| H | \geqq 1$. In the case, $H = 0$, considered above the numerical values of the two correction terms containing $J_{H+s}(z)$ and $J_{-H+s}(z)$ were of the same magnitude, for which reason both of them had to be taken into account.
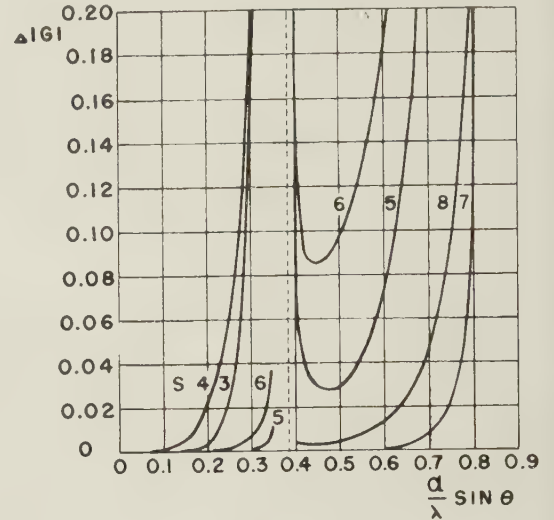
Fig. 4—The relative, maximum variation $\Delta| F_\theta |$ for a homogeneous ring array for $H = 0$.

However, for $H \geqq 1$ it will generally hold good that the correction term containing the function $J_{-H+s}(z)$ will be much larger than the term containing $J_{H+s}(z)$, the numerical value of a Bessel function with a constant argument decreasing fast though not necessarily monotonically with increasing order. Correspondingly, for $H \leqq -1$ it will generally hold good that $J_{H+s}(z)$ will be much larger than $J_{-H+s}(z)$. Thus we find the following approximate expressions for the normalized electric field strength;

$$F_\theta \leqq -\sin\theta[(-i)^H e^{iH\varphi} J_H(z) \\ + (-i)^{\mp H+s} e^{\mp i(\mp H+s)\varphi} J_{\mp H+s}(z)]$$

$$= \begin{cases} -\sin\theta(-i)^H e^{iH\varphi}[J_H(z) + (-1)^H(-i)^s e^{-is\varphi} J_{-H+s}(z)] \\ \hspace{5cm} \text{for} \quad H = 1 \\ -\sin\theta(-i)^H e^{iH\varphi}[J_H(z) + (-i)^s e^{is\varphi} J_{H+s}(z)] \\ \hspace{5cm} \text{for} \quad H = -1 \end{cases}$$

From the expression for the field derived here we see that the numerical value of the correction term is constant, whereas the phase difference between the principal term and the correction term will increase linearly with φ. The polar diagram of the numerical value of the field is an undulating line in the shape of a circle with one period between two neighboring antennas. The extreme values will occur in the directions that are $\pi/2$ behind the directions towards the antennas and in the directions halfway between these directions. The relative, maximum variation is found to be

$$| F_\theta | = \frac{| J_{\mp H+s}(z) |}{| J_H(z) |}.$$

This expression being valid for an even as well as for an odd number of antennas, it is seen that in the present case no systematic advantage can be derived from using an odd instead of an even number of elements as was the case for $H = 0$. The relative, maximum variation in the case of $H = 4$ is plotted in Fig. 5.

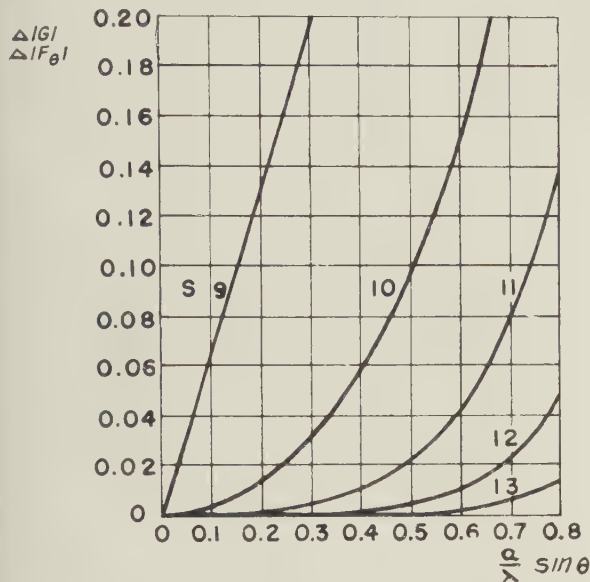
Fig. 5—The relative, maximum variation $\Delta |F_\theta|$ for a homogeneous ring array for $|H| = 4$.

Radiation resistance. For a homogeneous ring array of infinitely many Hertz dipoles Page has carried out a calculation of the radiation resistance.[18] From the derived expression it appears that the radiation resistance increases with increasing radius a of the ring array, and that it decreases with increasing increment of the current phase per revolution; i.e., with increasing H.

Directional Ring Array

Stenzel has shown that a ring array may be utilized for obtaining an antenna system with pronounced directional properties.[19] By giving the currents of the antennas in the ring a phase distribution such that there is constructive interference of the waves emitted in an arbitrary direction in space, hereafter called the principal direction, an array characteristic is obtained having a principal lobe in this direction. Stenzel considers especially the case where the principal direction is horizontal. If the principal direction is rotated in the horizontal plane, the horizontal diagram of the array characteristic will also rotate and, as a consequence of the

approximate rotational symmetry of the ring aerial, with its shape almost unchanged. For the ring aerial mentioned here, Stenzel has investigated the discrepancy between the array characteristic in the case of a finite, even number of elements and the characteristic in the case of infinitely many elements, apparently assuming that there is no essential difference between the case of an even number of elements and that of an odd number of elements. This assumption is explicitly stated by Brückmann in his textbook on antennas in the chapter dealing with Stenzel's theory.[20] In the present paper it will be investigated whether the array characteristic of the directional ring aerial with an even number of antennas differs essentially from the characteristic of the ring aerial with an odd number of elements as was the case with the ring aerial investigated by Page.[14] This investigation is based on Stenzel's as well as on Page's theory.

Let s identical and identically oriented antennas be placed along a circle with a radius a as shown in Fig. 1, and let us introduce a spherical co-ordinate system as described above in connection with this figure. The current I_j in the jth antenna is expressed by

$$I_j = Ie^{i\delta j},$$

where I is a constant, whereas δ_j denotes the current phase of the jth antenna. Choosing as a reference antenna, an antenna placed at the center of the circle and carrying a current sI we find by using the conventional expression for the array characteristic $G(\theta, \varphi)$ of an antenna array the following expression for the array characteristic of the ring array considered here;

$$G = \frac{1}{s} \sum_{j=1}^{s} e^{i[\delta_j - ka \sin \theta \cos (\phi - u_j)]},$$

where, as above, $k = \omega\sqrt{\mu\epsilon}$ denotes the intrinsic propagation constant of space.

In accordance with Stenzel's suggestion,[19] the currents in the s antennas are now given such phases that the waves emitted in an arbitrarily chosen direction (θ_o, φ_o), the principal direction, are in phase. This is obtained by choosing

$$\delta_j = ka \sin \theta_o \cos (\varphi_o - u_j).$$

By this choice of the phases δ_j the expression for the array characteristics will be

$$G = \frac{1}{s} \sum_{j=1}^{s} e^{ika[\sin \theta_o \cos (\phi_o - u_j) - \sin \theta \cos (\phi - u_j)]}.$$

Following Stenzel in defining an angle ξ by the equation

$$\cos \xi = \frac{\sin \theta \cos \varphi - \sin \theta_o \cos \varphi_o}{\sqrt{(\sin \theta \cos \varphi - \sin \theta_o \cos \varphi_o)^2 + (\sin \theta \sin \varphi - \sin \theta_o \sin \varphi_o)^2}},$$

and a parameter ρ by the equation

$$\rho = a\sqrt{(\sin \theta \cos \varphi - \sin \theta_o \cos \varphi_o)^2 + (\sin \theta \sin \varphi - \sin \theta_o \sin \varphi_o)^2},$$

[18] H. Page, "Radiation resistance of ring aerials," *Wireless Engr.*, vol. 25, pp. 102–109; 1948.
[19] H. Stenzel, "Über die Richtkarakteristik von in einer Ebene angeordneten Strahlern," *ENT*, vol. 6, pp. 165–181; 1929.

[20] H. Brückmann, "Antennen. Ihre Theorie und Technik,' Leipzig, p. 113; 1939.

we may now express the array characteristic in the following way,

$$G = \frac{1}{s} \sum_{j=1}^{s} e^{-ik\rho \cos (\xi - u_j)}.$$

This expression is of a type similar to the expression derived above for the array characteristic for a homogeneous ring array with antenna currents having the same phase; i.e., with $H = 0$. Using a similar procedure to that for the homogeneous ring array, we may write the expression derived here

$$G = \sum_{q=0}^{\infty} (2 - \delta_{on})(-i)^{qs} J_{qs}(k\rho) \cos qs\xi.$$

If the number s of elements is infinitely large, the array characteristic will be expressed by the first term in this infinite series

$$G = J_o(k\rho).$$

When the number of elements is finite, the remaining terms in the series will express the deviation of the array characteristic from the array characteristic applying in the case of an infinite number of elements.

The array characteristic in the case of a finite number of elements may be transformed with advantage so that it clearly appears which terms are in phase with (or in phase opposition to), and which are in phase quadrature with, the principal term. In so doing we must distinguish between the cases where the number s of elements is even and those where it is odd.

s even.

$$G = J_o(k\rho) + 2 \sum_{q=1}^{\infty} J_{qs}(k\rho) \cos \left(\frac{\pi}{2} - \xi\right)qs.$$

s odd.

$$G = J_o(k\rho) + 2 \sum_{q=1}^{\infty} J_{2qs}(k\rho) \cos \left(\frac{\pi}{2} - \xi\right)2qs$$

$$- i2 \sum_{q=0}^{\infty} J_{(2q+1)s}(k\rho) \sin \left(\frac{\pi}{2} - \xi\right)(2q + 1)s.$$

The expression for the array characteristic for s even was derived by Stenzel approximately in the way described here. The correction terms in the case of s even are seen to be in phase with the principal term. On the other hand, in the case of s odd, there occur correction terms in phase with, as well as correction terms in phase quadrature with, the principal term. Let us for a moment suppose that we have chosen s so large that the correction terms are small compared with the maximum value of the principal term. The array characteristic may then be expressed sufficiently accurately by the principal term and the first term in each of the infinite series of correction terms.

s even.

$$G \cong J_o(k\rho) + 2J_s(k\rho) \cos \left(\frac{\pi}{2} - \xi\right)s.$$

s odd.

$$G \cong J_o(k\rho) - i2J_s(k\rho) \sin \left(\frac{\pi}{2} - \xi\right)s.$$
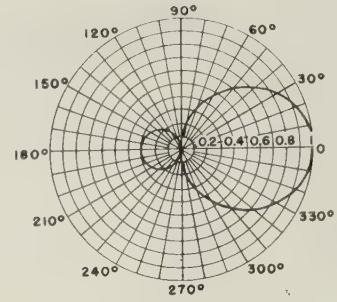
a. $\frac{a}{\lambda} = 0.25$.
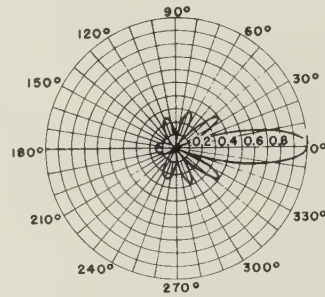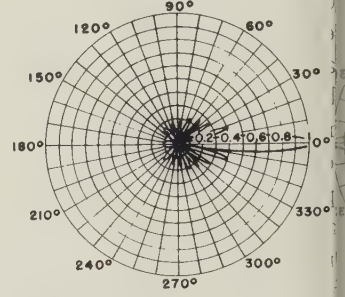
b. $\frac{a}{\lambda} = 1$.　　　　c. $\frac{a}{\lambda} = 2$.

Fig. 6—Horizontal diagram of the array characteristic for a directional ring array for various values of a/λ.

As the most important correction term in the case of s even is in phase with the principal term, whereas the most important correction term in the case of s odd is in phase quadrature with the principal term, the latter case must be assumed to be more favourable than the former in regard to the approximation of the array characteristic to the array characteristic valid for $s = \infty$. This will be illustrated later by a numerical example.

When the principal direction is vertical; i.e., when $\theta_o = 0$, the array will constitute a homogeneous ring array with currents in the same phase. The expression for the array characteristic derived here will then be reduced to the expression derived above for the array characteristic for a homogeneous ring array with $H = 0$. In what follows we shall illustrate the expression for the array characteristic by investigating a ring array with its principal direction in the horizontal plane.

For a ring array with a horizontal principal direction; i.e., with $\theta_o = \pi/2$, we find, confining the investigation to the field in the horizontal plane $\theta = \pi/2$,

$$\xi = \frac{\pi}{2} + \frac{\varphi + \varphi_o}{2},$$

$$\rho = 2a \sin \frac{\varphi - \varphi_o}{2}.$$

For a ring array with infinitely many elements and with the quotient between the radius and the wavelength $a/\lambda = 0.25$, 1, and 2 we hereby find the horizontal patterns of the array characteristic shown in Fig. 6(a)–6(c). By rotating the distribution of current phases

ese horizontal patterns will make a corresponding
tation in the horizontal plane. If the number of elements
infinitely large, this rotation will take place without
y deformation of the horizontal pattern of the array
aracteristic. On the other hand, if the number of
ments is finite, the radiation pattern will vary period-
lly, when the distribution of current phases is rotated.
On the basis of the exact expression for the array char-
teristic, the array characteristic has been calculated
a ring array with $a/\lambda = 0.5$ and with respectively
6, 7, and 8 elements for various values of the azimuth
of the principal direction. The horizontal patterns
und in this way are plotted in Fig. 7(a)–7(h). In this
ure the principal direction is fixed, whereas the an-
nna system rotates. The figure shows that for a small
mber of antennas, $s = 5$ or 6, the radiation pattern
viates considerably from the pattern valid for $s = \infty$.
owever, it is seen that 5 antennas give almost as good
approximation to the ideal pattern as 6, and 7 an-
nnas almost as good an approximation as 8. This
nfirms the assumption that an odd number of antennas
preferable to an even number. The advantage of using
odd number is, however, not so great here as in the
se of the homogeneous ring array dealt with above.

HOMOGENEOUS RING QUASI-ARRAY OF
TANGENTIAL ANTENNAS

omogeneous Ring Quasi-Array of Tangential Hertz
ipoles

Radiated Field; In a preceding section we have seen
at by using a homogeneous ring array of axial Hertz
poles we can obtain an azimuthally omnidirectional,
orizontally polarized field that is concentrated around
e horizontal plane. During recent years FM and
elevision have aroused an interest in antennas that
diate horizontally polarized fields. It is therefore an
vious thing in analogy to the investigation made
bove of the field from a ring array of axial dipoles to
vestigate the field from a ring quasi-array of equi-
istant, tangential Hertz dipoles carrying currents of
e same numerical value and with a phase that in-
eases uniformly along the circle. A ring array of this
nd will be called a homogeneous ring quasi-array. In
alogy with what was done in the case of the homo-
eneous ring array the increment of the current phase
uring one revolution will be denoted by $H2\pi$, where H
an integer.

As mentioned above, Carter has carried through a
lculation of the field from a ring array of a finite
umber of axial dipoles and from a ring quasi-array
a finite number of tangential or radial dipoles placed
oncentrically around a conducting cylinder and with
= 0 or $H = 1$.[15] In the case where the radius of the
nducting cylinder is equal to zero, and where the
rrent phases are characterized by $H = 0$ or $H = 1$,
e antenna systems treated by Carter constitute
pecial cases of the systems investigated here.

The field from an arbitrarily large, circular frame
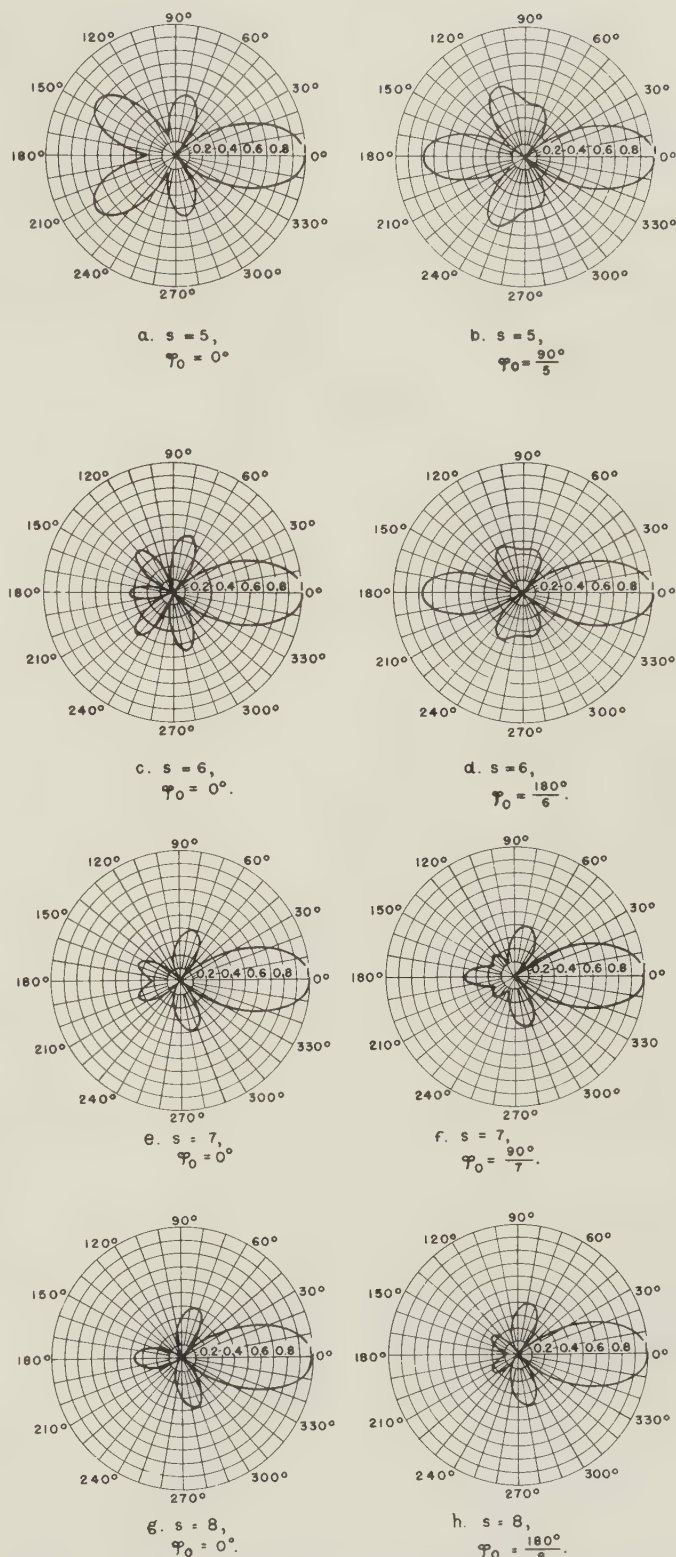
Fig. 7—Horizontal diagram of the array characteristic for a
directional ring array with $a/\lambda = 0.5$ for various values of the num-
ber s of elements.

aerial with constant current (i.e., $H = 0$) has previously
been investigated by Foster[21] and Moullin.[22] This field
calculation is included in this paper as a special case.

[21] D. Foster, "Loop antennas with uniform current," PROC.
IRE, vol. 32, pp. 603–607; October, 1944.
[22] E. B. Moullin, "Radiation from large circular loops" *Jour.
I. E. E.*, Pt. III, vol. 93, pp. 345–351; 1946.

The homogeneous ring quasi-array considered here is assumed to consist of s identical and identically oriented Hertz dipoles having the length L and being placed equidistantly along a circle with radius a tangential to this circle, as shown in Fig. 8. A spherical
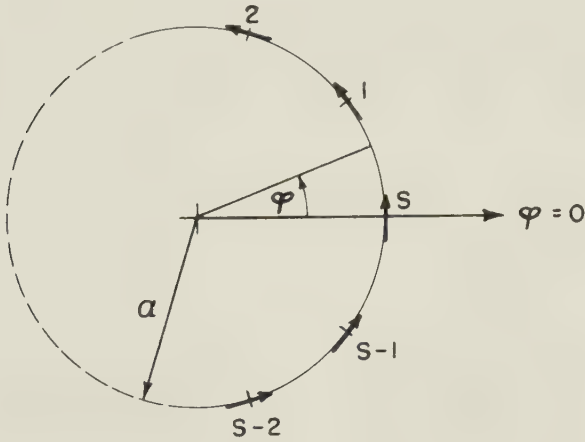
Fig. 8—Ring quasi-array of s tangential dipoles.

co-ordinate system (r, θ, φ) is placed as shown in the figure in the same way as the co-ordinate system shown in Fig. 1 so that the co-ordinates of antenna j are given by $(a, \pi/2, u_j)$, where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots, s.$$

It is assumed that the current I_j in the jth dipole is given by

$$I_j = I e^{iH u_j},$$

where I is a constant and H an integer, as mentioned above. We express the electric field strength $\overline{E}$ at the point (r, θ, φ) by using the notation introduced above,

$$\overline{E}(r, \theta, \varphi) = K \overline{F}(\theta, \varphi) \frac{e^{ikr}}{r},$$

where K is a normalization constant and $\overline{F}$ the dimension-free, normalized electric field strength. As in the investigation of the homogeneous ring array we choose here the normalization constant

$$K = \frac{iksL\zeta I}{4\pi}.$$

In a rectangular co-ordinate system placed in the usual way in relation to the spherical co-ordinate system we find by conventional methods the following expression for the components of the normalized electric field strength;

$$F_x = -\sum_{j=1}^{s} \frac{1}{s} e^{i[H u_j - ka(\cos u_j \sin \theta \cos \varphi + \sin u_j \sin \theta \sin \varphi)]} \sin u_j,$$

$$F_y = \sum_{j=1}^{s} \frac{1}{s} e^{i[H u_j - ka(\cos u_j \sin \theta \cos \varphi + \sin u_j \sin \theta \sin \varphi)]} \cos u_j,$$

$$F_z = 0,$$

or, by using the short-hand notation used above

$$F_x = -\frac{1}{s} \sum_{j=1}^{s} e^{i[H u_j - z \cos(\varphi - u_j)]} \sin u_j,$$

$$F_y = \frac{1}{s} \sum_{j=1}^{s} e^{i[H u_j - z \cos(\varphi - u_j)]} \cos u_j,$$

$$F_z = 0.$$

By using the series development given above,

$$e^{iz \cos v} = \sum_{n=0}^{\infty} (2 - \delta_{on}) i^n J_n(z) \cos nv,$$

we may express the rectangular components of $\overline{F}$ by

$$F_x = -\frac{1}{s} \sum_{j=1}^{s} e^{iH u_j} \sin u_j \sum_{n=0}^{\infty} (2 - \delta_{on})(-i)^n J_n(z)$$
$$\cos n(\varphi - u$$

$$= \frac{1}{4s} \sum_{n=0}^{\infty} (2 - \delta_{on})(-i)^{n+1} J_n(z) \sum_{j=1}^{s} \left\{ e^{in\varphi} \left[-e^{i(H-n+1) u_j} \right. \right.$$
$$\left. + e^{i(H-n-1) u_j} \right] + e^{-in\varphi} \left[-e^{i(H+n+1) u_j} + e^{i(H+n-1) u_j} \right] \right\},$$

$$F_y = \frac{1}{s} \sum_{j=1}^{s} e^{iH u_j} \cos u_j \sum_{n=0}^{\infty} (2 - \delta_{on})(-i)^n J_n(z)$$
$$\cos n(\varphi - u$$

$$= \frac{1}{4s} \sum_{n=0}^{\infty} (2 - \delta_{on})(-i)^n J_n(z)$$

$$\sum_{j=1}^{s} \left\{ e^{in\varphi} \left[e^{i(H-n+1) u_j} + e^{i(H-n-1) u_j} \right] \right.$$
$$\left. + e^{-in\varphi} \left[e^{i(H+n+1) u_j} + e^{i(H+n-1) u_j} \right] \right\},$$

$$F_z = 0.$$

The summations with respect to j in the above expressions are all of the type previously described;

$$\frac{1}{s} \sum_{j=1}^{s} e^{im u_j} = \frac{1}{s} \sum_{j=1}^{s} e^{\frac{i2\pi mj}{s}}$$

$$= \begin{cases} 1 & \text{for } m/s = p, \text{ wher} \\ & p = 0, \pm 1, \pm 2, \cdots \\ 0 & \text{in any other case,} \end{cases}$$

where m denotes an integer. By using this formula together with the following recursion formulas for Bessel functions,

$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x} J_n(x),$$

$$J_{n-1}(x) - J_{n+1}(x) = 2 J_n'(x),$$

and by assuming that

$$s > |H| + 1,$$

we find, after some manipulation,

$$F_x = \frac{1}{2} \sum_{q=0}^{\infty} (2 - \delta_{oq}) \left\{ \left[\frac{H + qs}{z} J_{H+qs}(z) \cos \varphi \right. \right.$$
$$\left. - i J'_{H+qs}(z) \sin \right.$$
$$e^{i(H+qs)(\varphi - \pi/2)} + \left[-\frac{-H + qs}{z} J_{-H+qs}(z) \cos \varphi \right.$$
$$\left. - i J'_{-H+qs}(z) \sin \varphi \right] e^{-i(-H + qs)(\varphi + \pi/2)} \right\},$$

$$F_y = \frac{1}{2} \sum_{q=0}^{\infty} (2 - \delta_{oq}) \left\{ \left[i J'_{H+qs}(z) \cos \varphi \right. \right.$$

$$+ \frac{H + qs}{z} J_{H+qs}(z) \sin \varphi]$$

$$e^{i(H+qs)(\varphi - \pi/2)} + [iJ'_{-H+qs}(z) \cos \varphi$$

$$- \frac{-H + qs}{z} J_{-H+qs}(z) \sin \varphi]$$

$$e^{-i(-H+qs)(\varphi + \pi/2)} \},$$

$$F_z = 0.$$

From the expressions for the rectangular components of the normalized electric field strength obtained here we find the following expressions for its spherical components:

$$F_\theta = F_x \cos \theta \cos \varphi + F_y \cos \theta \sin \varphi - F_z \sin \theta$$

$$= \frac{1}{2} \cos \theta \sum_{q=o}^\infty (2 - \delta_{oq}) \left[\frac{H + qs}{z} J_{H+qs}(z) e^{i(H+qs)(\phi - \pi/2)} \right.$$

$$\left. - \frac{-H + qs}{z} J_{-H+qs}(z) e^{-i(-H+qs)(\phi + \pi/2)} \right],$$

$$F_\varphi = -F_x \sin \varphi + F_y \cos \varphi$$

$$= \frac{i}{2} \sum_{q=o}^\infty (2 - \delta_{oq}) [J'_{H+qs}(z) e^{i(H+qs)(\phi - \pi/2)}$$

$$+ J'_{-H+qs}(z) e^{-i(-H+qs)(\phi + \pi/2)}].$$

Let us first consider the case where the number of elements is infinitely large. Any other term in the above expressions except those corresponding to $q = o$ will then be zero, and we find

$$F_\theta = \frac{H}{z} J_H(z) \cos \theta e^{iH(\phi - \pi/2)},$$

$$F_\varphi = iJ'_H(z) e^{iH(\phi - \pi/2)}.$$

These expressions show that in general both the θ- and the φ-components of the electric field strength will be different from zero. Further, as they are in phase quadrature with each other, the radiated field will in general be elliptically polarized. Only in the case of $H = 0$; i.e., for a circular frame aerial with a constant current, the θ-component of the electric field strength is zero in any direction, so that the field will be horizontally polarized everywhere. For $|H| \geqq 1$, in the horizontal plane, points having the same phase will be situated on an Archimedes' spiral as shown in Fig. 2 for the case $H = \pm 1$. The numerical values of the field components will be independent of φ.

In particular we shall consider the case, where the radius a of the ring quasi-array is much smaller than the wavelength λ divided by 2π; i.e., where $ka \ll 1$. The fact is that it is appropriate to let the radius of the ring quasi-array have so small a value that the polar diagram of its radiated field deviates little from the polar diagram occurring in the case of an infinitely small radius. The Bessel functions in the expressions derived above may then approximately be replaced by the first term of their series development, so that we find

$$F_\theta = \begin{cases} 0 & \text{for } H = 0, \\ \dfrac{(ka)^{H-1}}{2^H (H - 1)!} \cos \theta \sin^{H-1} \theta e^{iH(\phi - \pi/2)} & \text{for } |H| \geqq 1, \end{cases}$$

$$F_\varphi = \begin{cases} -i \dfrac{ka}{2} \sin \theta & \text{for } H = 0, \\ i \dfrac{(ka)^{H-1}}{2^H (H - 1)!} \sin^{H-1} \theta e^{iH(\phi - \pi/2)} & \text{for } |H| \geqq 1. \end{cases}$$

The numerical values of F_θ and F_φ are plotted in Fig. 9(a) and $(c) - (e)$ as functions of θ for various values of H. For any value of H the field will be horizontally polarized in horizontal directions. For $H = 0$ we obtain the well-known polar diagram of the radiated field from a small frame aerial. In the case where $|H| = 1$, and in this case only, the radiation in the direction of the axis is different from zero; in this case the field will be circularly

a. TANG. DIPOLES, H=0.
b. RAD. DIPOLES, H=0.
c. TANG. AND RAD. DIPOLES, H=1.
d. TANG. AND RAD. DIPOLES, H=2.
e. TANG. AND RAD. DIPOLES, H=9.
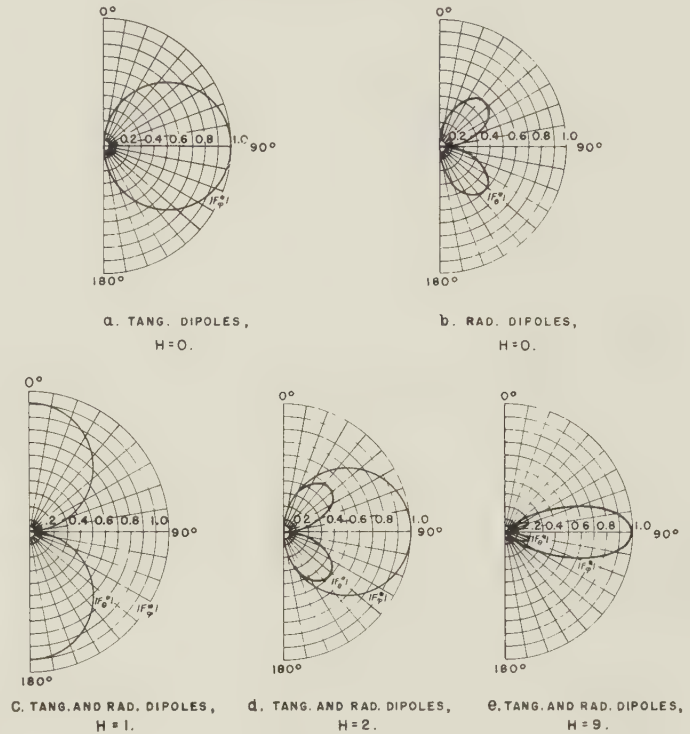
Fig. 9—Polar diagram of the electric field strength for a small homogeneous ring quasi array of tangential or radial dipoles for various values of H.

polarized in this direction. For $|H| \geqq 2$ the field will have the same character for any value of H; however, the field concentrates to an increasing degree around the horizontal plane for increasing $|H|$. The quotient between the maximum values of the θ- and the φ-components of the electric field strength is expressed approximately by

$$\frac{|F_\theta|_{max}}{|F_\varphi|_{max}} \simeq \frac{1}{\sqrt{e(|H| - 1)}},$$

where e is the base of the natural logarithms. Accordingly, with increasing $|H|$ the field will to an increasing degree become purely horizontally polarized in any direction, as is also seen from Fig. 9.

With an increasing radius of the ring quasi-array the field will to an increasing degree deviate from the field occurring in the case of an infinitely small radius of the

quasi-array. However, even in the case of a radius so large as the minimum radius that can be chosen in order to avoid super-gain, the shape of the field will deviate little from the shape occurring in the case of an infinitely small radius.

In the above calculations we have assumed that the number of elements is infinitely large; the ring quasi-array was then exactly azimuthally omnidirectional. Apart from the case of $H = 0$ and a very small radius, i.e., a small circular frame aerial with a constant current, the ring quasi-array must, however, in practice be composed of a finite number of elements. In order that the field from such a homogeneous ring quasi-array with a finite number of elements may be a fairly good approximation to the field from the corresponding quasi-array with infinitely many elements, the terms corresponding to $q = 0$ in the expressions for F_θ and F_φ must be the principal terms in the infinite series. The approximate condition that this will occur is that

$$s > 2\,|\,H\,|.$$

Assuming that this inequality is satisfied, we may generally express the field with a sufficiently good approximation by including the first correction term. So doing we find

$$F_\theta \cong \cos\theta\Big\{(-i)^H\frac{H}{z}J_H(z)$$
$$+ (-i)^s\Big[(-i)^H e^{is\phi}\frac{H+s}{z}J_{H+s}(z)$$
$$- (-i)^{-H}e^{-is\phi}\frac{-H+s}{z}J_{-H+s}(z)\Big]\Big\}e^{iH\phi},$$
$$F_\phi \cong i\Big\{(-i)^H J'_H(z) + (-i)^s\Big[(-i)^H e^{is\phi}J'_{H+s}(z)$$
$$+ (-i)^{-H}e^{-is\phi}J'_{-H+s}(z)\Big]\Big\}e^{iH\theta}.$$

In analogy with what we did in the case of the homogeneous ring array we shall here express the irregularity introduced into the field as a consequence of the fact that the number of elements is finite, by stating the relative, maximum variation of F_θ and F_φ. These relative, maximum variations are calculated on the basis of the expressions derived here analogously to the variations in the case of the homogeneous ring array. The description of this calculation is omitted here. The relative, maximum variation $\Delta\,|\,F_\theta\,|$ of the θ-component of the electric field strength for $|\,H\,| = 4$ is plotted in Fig. 10, and the relative, maximum variation $\Delta\,|\,F_\varphi\,|$ of the φ-component of the electric field strength for $H = 0$, and $|\,H\,| = 4$ is plotted in Fig. 11(a) and 11(b). Also in the case of a homogeneous ring quasi-array of tangential dipoles the case of $H = 0$ is seen to exhibit the characteristic feature that 3 antennas give a better polar diagram than 4, and 5 antennas a better diagram than 6, etc.

Radiation Resistance and Gain

We shall here calculate the radiation resistance and the gain of a homogeneous ring quasi-array of tangential dipoles by using the Poynting vector method. For the sake of simplicity we shall confine ourselves to dealing

Fig. 10—The relative, maximum variation of one of the field components for a homogeneous ring quasi-array of tangential dipoles ($\Delta\,|\,F_\theta\,|$) or radial dipoles ($\Delta\,|\,F_\phi\,|$) for $|\,H\,| = 4$.

with the case where the number of dipoles is infinitely large; in practice, the result obtained in this way will apply with a good approximation also when the number of dipoles is finite.

Denoting the intrinsic impedance of space by $\xi = \sqrt{\frac{\mu}{\epsilon}}$, we may express the Poynting vector $\bar{S}$ for an antenna with the normalization constant K and the normalized electric field strength $\bar{F}(\theta, \varphi)$ by

$$\bar{S} = \frac{\eta}{2}\,|\,\bar{E}\,|^2\hat{r} = \frac{\eta\,|\,K\,|^2}{2r^2}\,m(\theta, \varphi)\hat{r},$$

where $\hat{r}$ denotes a unit vector pointing in the direction of the r-axis, and where the function $m(\theta, \varphi)$ is defined by

$$m(\theta, \varphi) = |\,F_\theta\,|^2 + |\,F_\varphi\,|^2.$$

The effect P radiated through a spherical surface having its center at the origin and the radius r, is expressed by

$$P = \int_0^\pi \int_0^{2\pi} \mathrm{Re}(\bar{S}\cdot\hat{r})r^2\sin\theta\,d\theta\,d\varphi = 2\pi\eta\,|\,K\,|^2 M,$$

where the constant M is defined by

$$M = \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} m(\theta, \varphi)\sin\theta d\theta d\varphi.$$

If $|\,F_\theta\,|$ and $|\,F_\varphi\,|$ and consequently also $m(\theta, \varphi)$ are independent of φ, as is the case with the ring quasi-array we are considering here, we have

$$M = \frac{1}{2}\int_0^\pi m(\theta, \varphi)\sin\theta d\theta.$$

The radiation resistance R is defined as the resistance through which a certain reference current I' must flow in order that an effect equal to the radiated effect P may be developed,

$$P = \tfrac{1}{2}R\,|\,I'\,|^2.$$
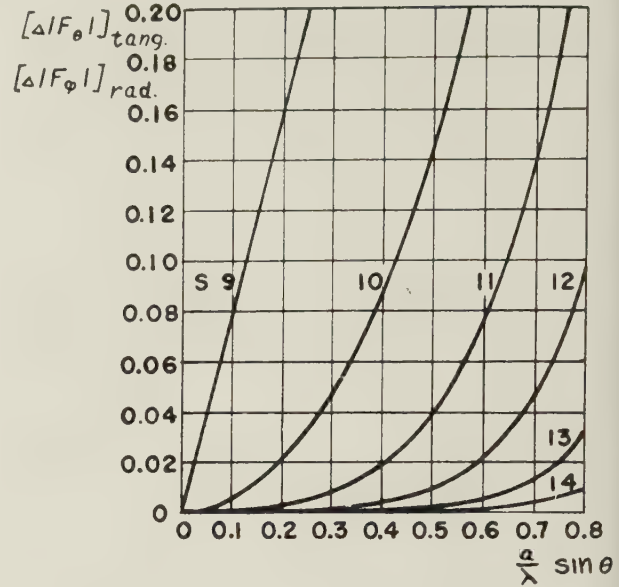
Consequently we find

$$R = NM,$$

a. $H = 0$.

b. $|H| = 4$.
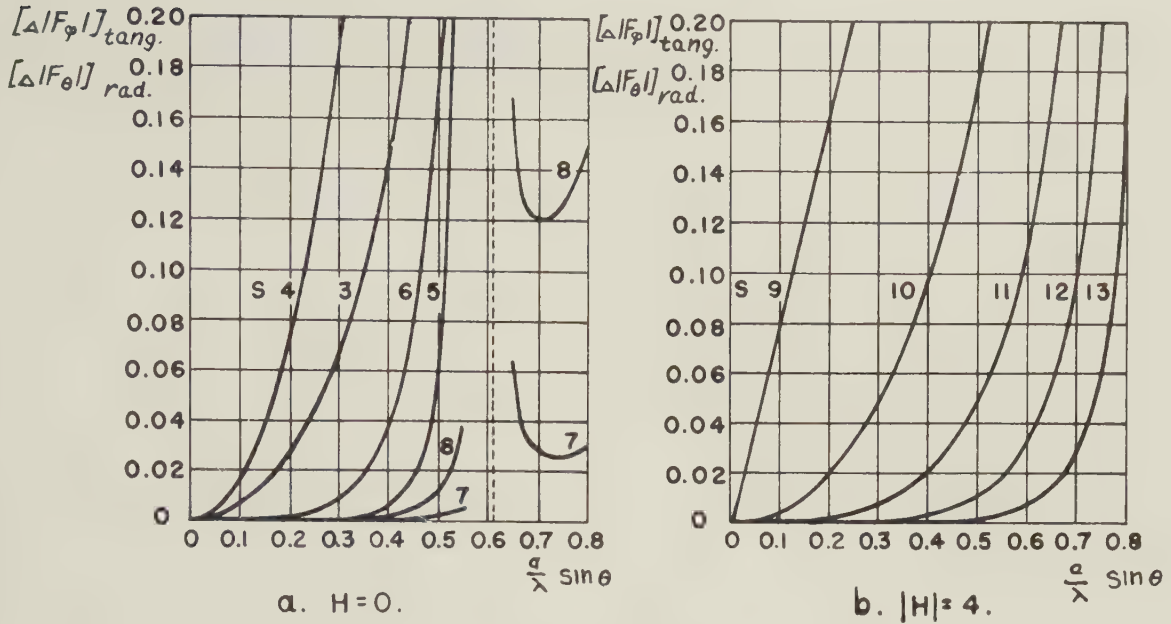
Fig. 11—The relative, maximum variation of one of the field components for a homogeneous ring quasi-array of tangential dipoles $(\Delta \mid F_\phi \mid)$ or radial dipoles $(\Delta \mid F_\theta \mid)$ for (a) $H = 0$ and (b) $\mid H \mid = 4$.

where

$$N = \frac{4\pi\eta \mid K \mid^2}{\mid I' \mid^2}.$$

The gain $g(\theta, \varphi)$ of the antenna in a certain direction (θ, φ) may be expressed by the parameters defined above, the function $m(\theta, \varphi)$ and the constant M. We find

$$g(\theta, \varphi) = \frac{\mid \bar{S} \mid}{\mid \bar{S} \mid \text{ mean}} = \frac{m(\theta, \varphi)}{M}.$$

Accordingly, the radiation resistance R as well as the gain $g(\theta, \varphi)$ of an antenna are determined by the constants N and M and the function $m(\theta, \varphi)$.

As the reference current I' for the homogeneous ring quasi-array in question we shall use the current I in one of the elements, no matter which of them. The constant N is then defined by

$$N = \frac{(ksL)^2 \zeta}{4\pi}.$$

The function $m(\theta, \varphi)$ is expressed by

$$m(\theta, \varphi) = \left(\frac{H}{z} J_H(z) \cos \theta\right)^2 + \left(J'_H(z)\right)^2,$$

whereas the constant M calculated by the use of this expression is given by

$$M = \frac{1}{2} \int_0^\pi \left[\left(\frac{H}{z} J_H(z) \cos \theta\right)^2 + (J'_H(z))^2\right] \sin \theta d\theta.$$

It may be shown that this integral may be expressed by known, tabulated functions and by the function $\bar{J}_n(x) = \int_0^x J_n(t) dt$ for any value of H. In the case of $H = 0$ this was demonstrated already by Foster.[21] Consequently, we confine ourselves to carrying through the calculation for $\mid H \mid \geq 1$. Through a simple transformation we find

$$M = 1/8 \left[A_{H-1}(ka) + A_{H+1}(ka) - 2B_H(ka)\right],$$

where the functions $A_n(x)$ and $B_n(x)$ are defined by

$$A_n(x) = \int_0^\pi J_n^2(x \sin \theta)(2 - \sin^2\theta) \sin \theta d\theta,$$

$$B_n(x) = \int_0^\pi J_{n-1}(x \sin \theta) J_{n+1}(x \sin \theta) \sin^3\theta d\theta.$$

As shown in an earlier publication these integrals may be expressed in the following way[1]

$$A_n(x) = \left(\frac{3}{2x} - \frac{4n^2 - 1}{8x^3}\right) \bar{J}_{2n}(2x)$$

$$+ \frac{2n - 1}{4x^2} \bar{J}_{2n}(2x) - \frac{1}{2x} J_{2n+1}(2x),$$

$$B_n(x) = \left(-\frac{1}{2x} + \frac{3(4n^2 - 1)}{8x^3}\right) J_{2n}(2x)$$

$$- \frac{3(2n - 1)}{4x^2} J_{2n}(2x) + \frac{3}{2x} J_{2n+1}(2x).$$

For the case of $n = 0$ several tables of the function $\bar{J}_n(x)$ occurring here have been published; the most comprehensive of these tables seems to be the one calculated by Lowan and Abramowitz.[23] On the basis of this table the author has calculated a table of $J_n(x)$ for $n = 1, 2 \cdots, 8$ and for $x = 0 (0.01) 10$.[1] The values of the function are given to five places of decimals; there is an inaccuracy of ± 1 in the last decimal place.

From the expressions for N, $m(\theta, \varphi)$, and M obtained here we can now easily calculate the radiation resistance R and the gain $g(\theta, \varphi)$ of the homogeneous ring quasi-array of tangential dipoles. The radiation resistance R

[23] A. N. Lowan and M. Abramowitz, "Table of the integrals" $\int_0^x J_0(t) dt$ and $\int_0^x Y_0(t) dt$," Jour. Math. Phys., vol. 22, pp. 2–12; 1943.

Fig. 12—$(2\pi a/sL)^2$ multiplied by the radiation resistance of a homogeneous ring quasi-array of tangential dipoles as a function of a/λ for various values of H.

the azimuth and originating from the finite number of elements. As an example of such a field calculation we shall here give the expression for the field from a homogeneous ring quasi-array of linear wire antennas with a constant current. As a special case of this we shall obtain the expression for the field from a polygonal frame aerial with a constant current.

Fig. 13—The gain in horizontal and vertical directions for a homogeneous ring quasi-array of tangential dipoles as a function of a/λ for various values of H.

multiplied by $(2\pi a/sL)^2$ is plotted in Fig. 12 as a function of a/λ for various values of H. The radiation resistance is seen to decrease with increasing increment of the current phase and to increase, though not monotonically, with increasing radius. In Fig. 13 is plotted the gain $g(\theta, \varphi)$ as a function of a/λ and for various values of H, partly for horizontal directions $\theta = \pi/2$ and partly for vertical directions $\theta = 0$ or π. Only the ring quasi-array with $|H| = 1$ has in the vertical direction a gain that is different from zero.

Homogeneous Ring Quasi-Array of Tangential, Linear Antennas with a Constant Current.

In dealing with the homogeneous ring quasi-array of tangentially oriented dipoles we have so far assumed the elements to be Hertz dipoles; i.e., infinitely short dipoles. For many applications this was a good approximation, and it made the mathematical treatment rather simple so that in such cases where this approximation was permissible, the principally important facts were emphasized more than would have been the case if less idealizing assumptions had been made. In practice, however, a ring quasi-array of the type considered here must be composed of antennas of finite dimensions, e.g., of a number of linear wire antennas. For the simple case where all currents are in the same phase; i.e., for $H = 0$, such a design has been suggested by Moullin.[22] However, if the current distribution on the various antennas is known, it is possible to carry out a calculation of the radiated field along the same lines as we did above in calculating the field from a homogeneous ring quasi-array of Hertz dipoles. The field will then be expressed by a principal term that is independent of the azimuth, in addition to correction terms dependent on

We consider a ring quasi-array of s tangentially oriented, linear antennas having the same length l. The mid-points of the antennas are assumed to be placed equidistantly along the periphery of a circle with radius a, as shown in Fig. 14. The antenna j will thus coincide
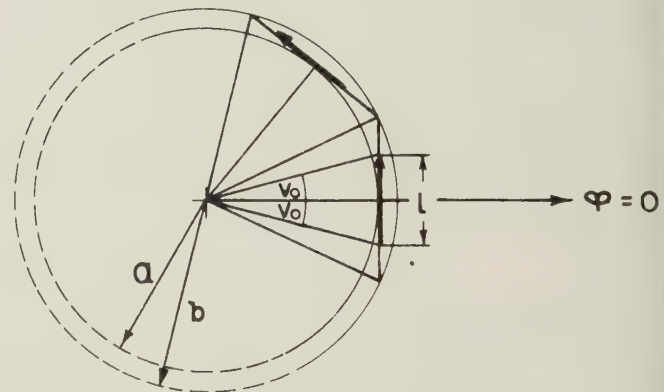
Fig. 14—Ring quasi-array of s tangentially oriented linear antennas of length l.

with the side j in a regular s-angle, the inscribed circle of which is the said circle with radius a. A spherical co-ordinate system is introduced as shown in the figure so that the mid-point of antenna j has the co-ordinates $(a, \pi/2, u_j)$, where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots, s.$$

The current I_j in antenna j is supposed to be given by

$$I_j = Ie^{iHu_j}.$$

analogy with the calculations previously carried out we introduce a normalizing constant K defined by

$$K = \frac{iksl\zeta I}{4\pi} \; ;$$

the field will then be determined by the specification of the normalized electric field strength $\bar{F}(\theta, \varphi)$ corresponding to this constant.

In order to be able to express F in a simple way we define the following functions;

$$P_{n,r}(x) = \frac{1}{tgr} \int_o^r J_n\left(\frac{x}{\cos v}\right) \frac{\cos nv}{\cos^2 v} \, dv,$$

$$Ps_{n,r}(x) = \tfrac{1}{2}[P_{n-1,r}(x) + P_{n+1,r}(x)],$$

$$Pd_{n,r}(x) = \tfrac{1}{2}[P_{n-1,r}(x) - P_{n+1,r}(x)].$$

By using these functions we may express the components of the normalized electric field strength in the following way

$$F_\theta = \tfrac{1}{2} \cos\theta \sum_{q=o}^\infty (2 - \delta_{oq})[Ps_{H+qs,v_o}(z)e^{i(H+qs)(\phi-\pi/2)}$$
$$- Ps_{-H+qs,v_o}(z)e^{-i(-H+qs)(\phi+\pi/2)}],$$

$$F_\phi = \frac{i}{2} \sum_{q=o}^\infty (2 - \delta_{oq})[Pd_{H+qs,v_o}(z)e^{i(H+qs)(\phi-\pi/2)}$$
$$+ Pd_{-H+qs,v_o}(z)e^{-i(-H+qs)(\phi+\pi/2)}],$$

where

$$z = ka \sin\theta$$

and

$$v_o = \text{arc } tg\,\frac{l}{2a} \, .$$

It is seen that this expression may be derived from the expression for the field from a homogeneous ring quasi-array of tangential Hertz dipoles given above by replacing $n/zJ_n(z)$ and $J'_n(x)$ by $Ps_{n,v_o}(x)$ and $Pd_{n,v_o}(x)$ respectively. When the length l of the antennas converges towards zero, the expressions for the field components derived here converge towards the expressions for the field components of a corresponding quasi-array of Hertz dipoles.

In the case where the number s of the elements is infinitely large we find from the expressions given above the following expressions for the field components,

$$F_\theta = \cos\theta Ps_{H,v_o}(z)e^{iH(\phi-\pi/2)},$$

$$F_\phi = iPd_{H,v_o}(z)e^{iH(\phi-\pi/2)}.$$

For a homogeneous ring quasi-array of antennas with a sine-shaped current distribution corresponding expressions for the field may be derived; the functions which must be introduced in this case are more complicated than those occurring here.

Homogeneous Ring Quasi-Array of Radial Antennas

Homogeneous Ring Quasi-Array of Radial Hertz Dipoles Radiated Field; In the preceding section we have seen that a homogeneous ring quasi-array of tangential antennas radiates an essentially horizontally polarized field which to an increasing degree concentrates around the horizontal plane with increasing increment of the current phase per revolution. It may be supposed that a homogeneous ring quasi-array of radial dipoles radiates a field of a similar type. We shall therefore carry out an investigation of the field from such an antenna system.

Certain types of homogeneous ring quasi-array have long since found practical application. Böhm has suggested an antenna with reduced radiation for high elevation angles, and which consists of a vertical wire being connected at its upper end to a system of radial wires of such a length that regard has to be taken of the radiation from these wires.[12] The radiation from the radial wires results in a reduction of the field radiated at high elevation angles so that the antenna may be used as a fading-reducing antenna. Böhm[24] has calculated the field radiated by the radial wires on the assumption that the number of wires is infinitely large, so that they form a disk, and so that the total current passing any circle concentric with the periphery of the disk is constant. Whereas this antenna is composed of radial wires with currents in the same phase, the turnstile antenna invented by Brown[25] may be considered constructed by radial wires with currents the phase of which increases 2π during one revolution. Thus, with the notation used above, these antennas correspond to $H = 0$ and $H = 1$ respectively.

As the simplest type of a homogeneous ring quasi-array of radial antennas we shall here consider such a ring quasi-array of radial Hertz dipoles. For the special values of H, $H = 0$ and $|H| = 1$, the field from this quasi-array was previously calculated by Carter.[15] We extend the investigation of the ring quasi-array in question to quasi-arrays with an arbitrary value of H, so that the possibility of arriving at new, usable antenna systems of this type is explored.

We shall calculate the field from the ring quasi-array shown in Fig. 15, of s radially oriented Hertz dipoles of the length L placed equidistantly along a circle with radius a. A spherical co-ordinate system (r, θ, φ) is introduced so that the co-ordinates of the jth dipole will be $(a, \pi/2, u_j)$, where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots, s.$$

The current I_j in the jth dipole is assumed to be

$$I_j = Ie^{iHu_j}.$$

The calculation of the field is carried out in the same manner as for a homogeneous ring quasi-array of tangential dipoles. Introducing the normalization constant K, defined by

$$K = \frac{iksL\zeta I}{4\pi},$$

[24] O. Böhm, "Rundfunk-Sendeantennen mit vertikal gebündelter Ausstrahlung," *Hochfreq. u. Elektroak.*, vol. 42, pp. 137–145; 1933.
[25] G. H. Brown, "Turnstile aerials," *Electronics*, vol. 9, pp. 14–17, 48; April, 1936.

we find the following expressions for the components of the normalized electric field strength $\bar{F}(\theta, \phi)$;

$$F_\theta = \frac{i}{2} \cos\theta \sum_{q=0}^{\infty} (2 - \delta_{oq})[J'_{H+qs}(z)e^{i(H+qs)(\phi-\pi/2)}$$
$$+ J'_{-H+qs}(z)e^{-i(-H+qs)(\phi+\pi/2)}],$$

$$F_\phi = -\frac{1}{2} \sum_{q=0}^{\infty} (2 - \delta_{oq})\left[\frac{H+qs}{z} J_{H+qs}(z)e^{i(H+qs)(\phi-\pi/2)} \right.$$
$$\left. - \frac{-H+qs}{z} J_{-H+qs}(z)e^{-i(-H+qs)(\phi+\pi/2)}\right],$$
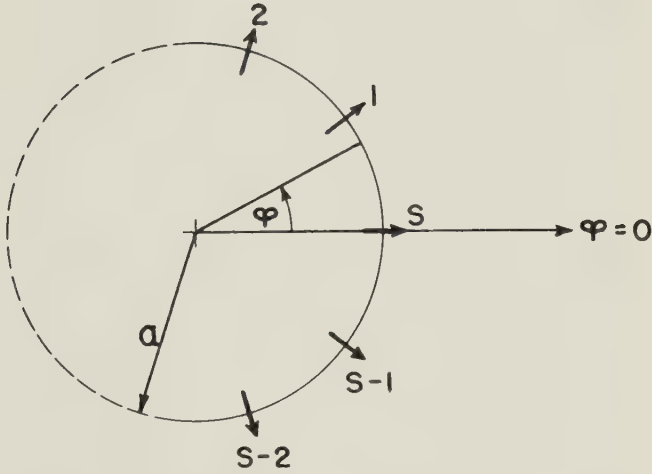
where

$$z = ka \sin\theta.$$

Fig. 15—Ring quasi-array of s radial dipoles.

We shall first consider the field in the case where the number of elements is infinitely large. From the above expression we find, for $s = \infty$,

$$F_\theta = iJ'_H(z) \cos\theta e^{iH(\phi-\pi/2)},$$

$$F_\varphi = -\frac{H}{z} J_H(z)e^{iH(\phi-\pi/2)}.$$

The radiated field will thus in general be elliptically polarized. Only for $H = 0$ the φ-component of the field strength is zero in every direction so that the field is polarized in the direction of the θ-axis. For $|H| \geqq 1$ points in the horizontal plane having the same phase will be situated on an Archimedes' spiral just as in the case of the homogeneous ring arrays and ring quasi-arrays considered earlier in this paper.

A general idea of the structure of the field for various values of H is best obtained by investigating the field for an infinitely small radius of the ring quasi-array. From the expressions given above we find for $ka \ll 1$

$$F_\theta = \begin{cases} -i\frac{ka}{2} \cos\theta \sin\theta & \text{for } H = 0, \\ i\frac{(ka)^{H-1}}{2^H(H-1)!} \cos\theta \sin^{H-1}\theta e^{iH(\phi-\pi/2)} & \text{for } |H| \geqq 1, \end{cases}$$

$$F_\varphi = \begin{cases} 0 & \text{for } H = 0, \\ -\frac{(ka)^{H-1}}{2^H(H-1)!} \sin^{H-1}\theta e^{iH(\phi-\pi/2)} & \text{for } |H| \geqq 1. \end{cases}$$

Fig. 9(b) shows the field for $H = 0$. In the case $|H| \geqq 1$ the field from the ring quasi-array considered here is seen to be identical with the field from the ring quasi-array of tangential Hertz dipoles previously investigated. For $|H| = 1, 2$, and 9 this field is plotted in Fig. 9(c)–9(e). The comments made on the ring quasi-array of tangential dipoles are thus applicable here too.

It is only for an infinitely small radius and for $|H| \geqq$ that the field from a homogeneous ring quasi-array of radial dipoles is identical with the field from a corresponding homogeneous ring quasi-array of tangential dipoles. With an increasing radius of the ring quasi-array of radial dipoles considered here the radiated field will deviate to an increasing degree from the ideal field shown in Fig. 9 as well as from the field radiated from a corresponding ring quasi-array of tangential dipoles.

So far we have assumed that the number of elements is infinitely large. In practice, however, the number will be finite, and we must then use the complete expression for the field given earlier. We suppose that the number of elements is larger than $2|H|$,

$$s > 2|H|.$$

The term corresponding to $q = 0$ in the expressions for F_θ and F_φ will then be the dominant term, and we may therefore express the field components by including only one extra term in each of the infinite series

$$F_\theta \cong i \cos\theta[J'_H(z)e^{iH(\phi-\pi/2)} + J'_{H+s}(z)e^{i(H+s)(\phi-\pi/2)}$$
$$+ J'_{-H+s}(z)e^{-i(-H+s)(\phi+\pi/2)}]$$

$$F_\phi \cong -\left[\frac{H}{z} J_H(z)e^{iH(\phi-\pi/2)} + \frac{H+s}{z} J_{H+s}(z)e^{i(H+s)(\phi-\pi/2)} \right.$$
$$\left. - \frac{-H+s}{z} J_{-H+s}(z)e^{-i(-H+s)(\phi+\pi/2)}\right].$$

It appears from these expressions that the relative maximum variation $\Delta|F_\theta|$ and $\Delta|F_\varphi|$ for the quasi-array considered here is equal to $\Delta|F_\varphi|$ and $\Delta|F_\theta|$ respectively for the corresponding homogeneous ring quasi-array of tangential dipoles previously dealt with. Consequently Fig. 11(a) and 11(b) show the relative maximum variation of the θ-component for a homogeneous ring quasi-array with radially oriented elements for $H = 0$ and $|H| = 4$, and Fig. 10 the relative maximum variation of the φ-component for such a array for $|H| = 4$.

Radiation Resistance and Gain

We shall now calculate the radiation resistance and the gain for a homogeneous ring quasi-array of infinitely many Hertz dipoles. To a good approximation the result hereby obtained will also be applicable to a ring quasi-array with a finite number of elements when the number of elements is as large as the considerations to be taken for obtaining a fair approximation to an azimuthally omnidirectional antenna system permit. The calculation is made in close analogy to the corresponding calculation for a homogeneous ring quasi-array of tangential Hertz dipoles.

It being assumed that the radiation resistance R is given by the current in an odd dipole as a reference current, the constant N previously introduced will be

$$N = \frac{(ksL)^2 \zeta}{4\pi}.$$

Hereby the function $m(\theta, \varphi)$ will be expressed by

$$m(\theta, \varphi) = (J'_H(z) \cos \theta)^2 + \left(\frac{H}{z} J_H(z)\right)^2,$$

whereas the constant M calculated from this expression is given by

$$M = \frac{1}{2} \int_0^\pi \left[(J'_H(z) \cos \theta)^2 + \left(\frac{H}{z} J_H(z)\right)^2\right] \sin \theta \, d\theta.$$

We shall here confine ourselves to giving the expression for the radiation resistance for $|H| \geqq 1$. It may then be shown that the expression for M may be transformed into

$$M = \frac{1}{8} [A_{H-1}(ka) + A_{H+1}(ka) + 2B_H(ka)],$$

where $A_n(x)$ and $B_n(x)$ are the functions defined in a preceding section. From N, $m(\theta, \varphi)$, and M we may now calculate the radiation resistance and the gain. The radiation resistance R multiplied by $(2\pi a/sL)^2$ is plotted in Fig. 16 as a function of a/λ for the various values of
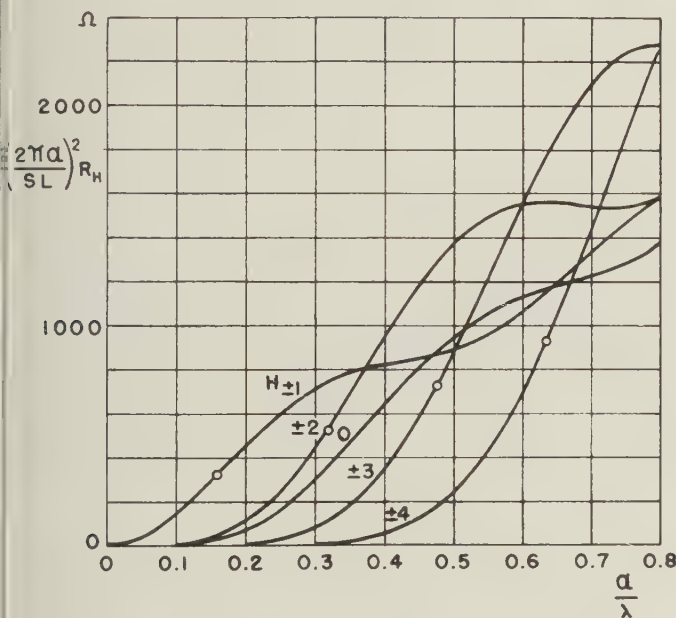
Fig. 16—$(2\pi a/sL)^2$ multiplied by the radiation resistance of a homogeneous ring quasi-array of radial dipoles as a function of a/λ for various values of H.

H, whereas the gain $g(\theta, \varphi)$ is plotted in Fig. 17. The curves showing the radiation resistance and the gain of the ring quasi-array investigated here are seen to have a shape similar to, but quantitatively deviating from, the corresponding curves for the ring quasi-array of tangentially oriented dipoles previously investigated.

Homogeneous Ring Quasi-Array of Radial, Linear Antennas with a Sine-Shaped Current

In the investigation made above of homogeneous ring quasi-arrays of radially oriented dipoles we have so far

assumed, for the matter of simplicity, that the elements of the ring quasi-arrays were Hertz dipoles. In practice, however, the quasi-arrays will be composed of linear antennas having a finite length. This applies to such antennas as the disk antenna suggested by Böhm[24] and the turnstile antenna suggested by Brown,[25] and it will apply also to ring quasi-arrays for larger values of $|H|$.
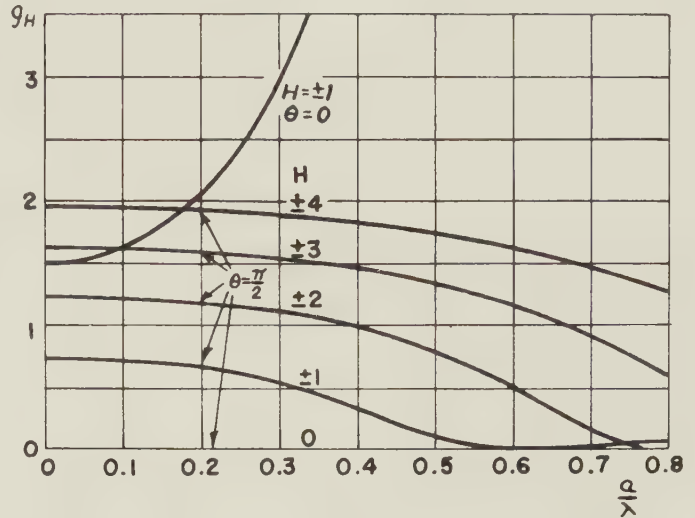
Fig. 17—The gain in horizontal and vertical directions for a homogeneous ring quasi-array of radial dipoles as a function of a/λ for various values of H.

When the current distribution on the linear antennas is known, it is possible to derive an expression for the field similar to the expression derived above for the field from a ring quasi-array of radial Hertz dipoles.
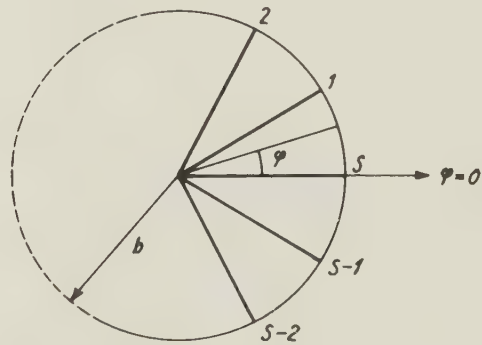
Fig. 18—Ring quasi-array of a radially oriented, linear antenna.

By way of example we shall here give the field from a ring quasi-array of s linear antennas of the length b, placed as the spokes of a wheel, as shown in Fig. 18, and carrying sine-shaped currents I_j of the same numerical value. The phase of these currents increases from one antenna to another so that its increment during one revolution amounts to $H2\pi$. With the notation introduced in Fig. 18 the current I_j in the current maximum (actual or virtual) of antenna j may be expressed by

$$I_j = Ie^{iHu_j},$$

where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots, s.$$

Introducing a spherical co-ordinate system (r, θ, φ) in the usual way and defining the normalization constant

$$K = \frac{iksb\zeta I}{4\pi},$$

we may express the components of the normalized electric field strength $\bar{F}(\theta, \varphi)$ by

$$F_\theta = \frac{i}{2} \frac{\cos\theta}{kb} \sum_{q=o}^{\infty} (2 - \delta_{oq})[pd_{H+qs,\theta}(kb)e^{i(H+qs)(\phi-\pi/2)}$$
$$+ pd_{-H+qs,\theta}(kb)e^{-i(-H+qs)(\phi+\pi/2)}],$$

$$F_\varphi = -\frac{1}{2} \frac{1}{kb} \sum_{q=o}^{\infty} (2 - \delta_{oq})[ps_{H+qs,\theta}(kb)e^{i(H+qs)(\phi-\pi/2)}$$
$$- ps_{-H+qs,\theta}(kb)e^{-i(-H+qs)(\phi+\pi/2)}.$$

The functions $ps_{n,\theta}(x)$ and $pd_{n,\theta}(x)$ occurring in these expressions are defined by

$$ps_{n,\theta}(x) = \tfrac{1}{2}[p_{n-1,\theta}(x) + p_{n+1,\theta}(x)],$$
$$pd_{n,\theta}(x) = \tfrac{1}{2}[p_{n-1,\theta}(x) - p_{n+1,\theta}(x)],$$

where

$$p_{n,\theta}(x) = \int_o^x \sin(x - t) J_n(t \sin\theta)\, dt.$$

It seems impossible generally to express this function in a simple way by known functions. However, an exception is the important case $\theta = \pi/2$ corresponding to radiation in the horizontal plane $\theta = \pi/2$. As shown by the author, we have

$$\int_o^x \sin(x - t) J_n(t)\, dt = \begin{cases} (-1)^{n/2}[n \sin x - x J_o(x) \\ \quad + 2 \sum_{j=1}^{\frac{n-1}{2}} (-1)^j(n + 1 - 2j) \\ \quad J_{2j-1}(x)] \quad \text{for } n \text{ odd,} \\ (-1)^{n/2}[n(\cos x - J_o(x)) \\ \quad + x J_1(x) + 2 \sum_{j=1}^{n/2-1} (-1)^{j+1} \\ \quad (n - 2j) J_{2j}(x)] \text{ for } n \text{ even.} \end{cases}$$

For $H = 1$, $s = 4$, and $b = \lambda/4$ the above formulas will express the field from an ordinary turnstile antenna.

Homogenous Ring Quasi-Array of Oblique Antennas

In the investigations made until now of homogeneous ring arrays and ring quasi-arrays of dipoles we have assumed that the dipoles were axial, tangential, or radial; we have hereby obtained knowledge of and a basis for the design of some useful types of antennas. However, it is conceivable that some types of ring quasi-arrays of obliquely oriented antennas may also turn out to be useful for practical purposes. An antenna system which in principle is a special case of such a ring quasi-array, is the antenna system invented by Lindenblad,[26] which has been further investigated by Brown and Woodward;[27] with the notation used here it may be

---

[26] N. E. Lindenblad, "Antennas and transmission lines at the Empire State Television Station. Part 2, "*Communications*, vol. 21, pp. 10–14, 24–26; April, 1941.

[27] G. H. Brown and O. M. Woodward, "Circularly-polarized omnidirectional antenna," *RCA Rev.*, vol. 8, pp. 259–269; 1947.

---

described as a homogeneous ring quasi-array, with $H = 0$, of linear antennas that are perpendicular to the radius at their center, and that form a suitable angle with the horizontal plane, this angle depending on the radius of the circle. This antenna system radiates in horizontal directions a circularly polarized field that is independent of azimuth; for this reason it is suited for communication between a fixed station and an airplane.

We shall here generalize the investigation made by Brown and Woodward, setting ourselves the task of calculating the field from a homogeneous ring quasi-array with an arbitrary H of an arbitrary number of dipoles with an arbitrary orientation. The antenna system treated by Brown and Woodward will then be comprised as a special case by this investigation, but the expression for the field will be presented here in a form more suitable as a basis for the design. The homogeneous ring arrays and ring quasi-arrays previously dealt with in this paper will also be included as special cases in the following investigation.
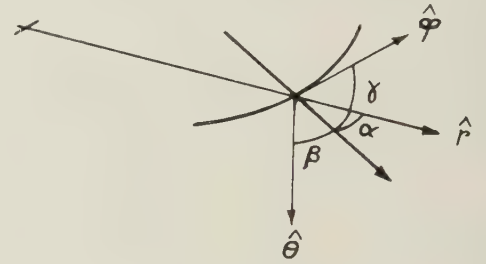
Fig. 19—Orientation of a dipole in a homogeneous ring quasi-array of oblique dipoles.

We consider s Hertz dipoles of the length L that are placed equidistantly along a circle with radius a, as shown in Fig. 1. The antennas are assumed to form the angles α, β, and γ with the r-, θ-, and φ-axes, respectively, in a spherical co-ordinate system, as shown in Fig. 19. The current in the jth dipole is supposed to be given by

$$I_j = Ie^{iHu_j},$$

where

$$u_j = \frac{2\pi j}{s} \qquad j = 1, 2, \cdots,$$

With the normalization constant K being defined by

$$K = \frac{iksL\zeta I}{4\pi},$$

the normalized electric field strength $\bar{F}(\theta, \varphi)$ may now be expressed by

$$F_\theta = \frac{1}{2} \sum_{q=o}^{\infty} (2 - \delta_{oq})\{[iJ'_{H+qs}(z) \cos\theta \cos\alpha$$
$$+ J_{H+qs}(z) \sin\theta \cos\beta$$
$$+ \frac{H+qs}{z} J_{H+qs}(z) \cos\theta \cos\gamma]e^{i(H+qs)(\varphi-\pi/2)}$$
$$+ [iJ'_{-H+qs}(z) \cos\theta \cos\beta$$

$$+ J_{-H+qs}(z) \sin\theta\cos\beta - \frac{-H+qs}{z} J_{-H+qs}(z)$$

$$\cos\theta\cos\gamma]e^{-i(-H+qs)(\varphi+\pi/2)}\Big\},$$

$$F_\varphi = \frac{1}{2}\sum_{q=0}^{\infty}(2-\delta_{oq})\Big\{\Big[-\frac{H+qs}{z}J_{H+qs}(z)\cos\alpha$$

$$+ iJ'_{H+qs}(z)\cos\gamma\Big]e^{i(H+qs)(\varphi-\pi/2)}$$

$$+\Big[\frac{-H+qs}{z}J_{-H+qs}(z)\cos\alpha + iJ'_{-H+qs}(z)\cos\gamma\Big]$$

$$e^{-i(-H+qs)(\varphi+\pi/2)}\Big\}.$$

The terms corresponding to $q = 0$ in these expressions denote the field corresponding to an infinitely large number of elements. Of the many possible antenna systems that may result from using different values of α, β, γ, and H we shall only give attention here to the following two important cases: $\alpha = \pi/2$, $H = 0$, and $\gamma = \pi/2$, $H = 1$.

$\alpha = \pi/2$, $H = 0$. In this case the dipoles are tangents to a circular cylinder. They form the angle γ with the horizontal plane, and they carry currents with the same phase. If we choose especially

$$\gamma = \pm \arctg \frac{J_1(ka)}{J_o(ka)},$$

the field components in the case of an infinite number of elements will be expressed by

$$F_\theta = \pm \frac{J_1(ka)}{J_o(ka)} J_o(z) \sin\theta\cos\gamma,$$

$$F_\varphi = -iJ_1(z)\cos\gamma.$$

The antenna system designed in that way is seen to be azimuthally omnidirectional and in horizontal directions to radiate a circularly polarized field. In the vertical direction the field is zero. The quasi-array dealt with here is the very antenna system invented by Lindenblad[26] and further investigated by Brown and Woodward[27] discussed previously.

When the ring quasi-array has such a small radius that $ka \ll 1$, we find

$$F_\theta \cong \pm \frac{ka}{2}\sin\theta,$$

$$F_\varphi \cong -i\frac{ka}{2}\sin\theta.$$

The field is then seen to be circularly polarized in any direction in space. The polar diagram for each of the field components becomes identical with the diagram shown in Fig. 3(a).

From the general expression given above for the field from a homogeneous ring quasi-array of an arbitrary number of oblique elements we easily find the relative, maximum variation of the antenna system considered here, when this system has a finite number of elements.

$\gamma = \pi/2$, $H = 1$. In this case the dipoles are placed along the edges of a pyramid; they make the angle α with the horizontal plane, and the current phases increase

2π during one revolution. The antenna system is azimuthally omnidirectional, and the field radiated in the vertical direction is circularly polarized. In horizontal directions the field will be linearly polarized. By choosing a suitable angle of inclination α for the antennas we may obtain that the field radiated in horizontal directions forms an angle of 45 degrees with the horizontal plane. We put

$$\alpha = \pm\arctg\frac{1}{ka},$$

and obtain hereby the following expressions for the field components in the case of an infinitely large number of antennas

$$F_\theta = \Big[iJ_1'(z)\cos\theta \pm \frac{1}{ka}J_1(z)\sin\theta\Big]\cos(\arctg\frac{1}{ka})e^{i(\varphi-\pi/2)},$$

$$F_\varphi = -\frac{1}{z}J_1(z)\cos\Big(\arctg\frac{1}{ka}\Big)e^{i(\varphi-\pi/2)}.$$

In the vertical direction this field is circularly polarized. In horizontal directions it is linearly polarized, and the electric field strength extends an angle of ± 45 degrees with the horizontal plane.

A general impression of the character of the field is best obtained by considering a ring quasi-array of the type considered here with such a small radius that $ka \ll 1$. We then have

$$F_\theta \cong \frac{1}{2}\cos\Big(\arctg\frac{1}{ka}\Big)[i\cos\theta \pm \sin^2\theta]e^{i(\varphi-\pi/2)},$$

$$F_\varphi \cong -\frac{1}{2}\cos\Big(\arctg\frac{1}{ka}\Big)e^{i(\varphi-\pi/2)}.$$
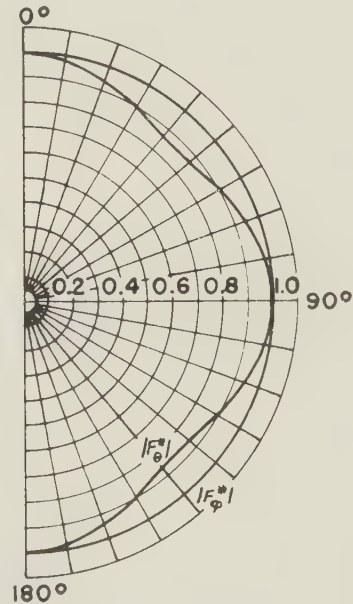
Fig. 20—The electric field radiated by a small homogeneous ring quasi-array of oblique dipoles with $\gamma = \pi/2$ and $H = 1$.

These field components are plotted in Fig. 20. For the polar distances $\theta = 0$ degrees, 30 degrees, 60 degrees, and 90 degrees the full-drawn line in Fig. 21 (a)–(d) shows the ellipse which is described by the vector, the electric field strength.

a. $\theta = 0°$

b. $\theta = 30°$

c. $\theta = 60°$

d. $\theta = 90°$
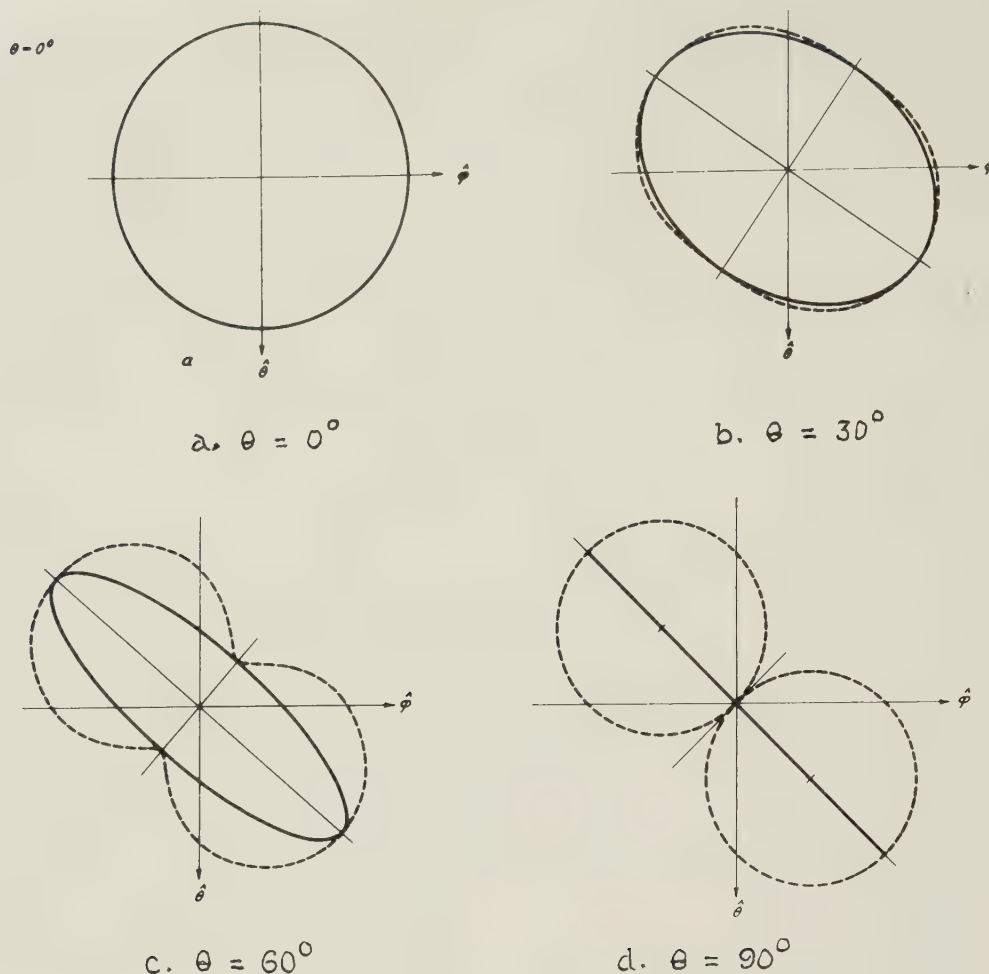
Fig. 21—The full-drawn lines show the polarization ellipses of the field radiated by a homogeneous ring quasi-array of oblique dipoles with $\gamma = \pi/2$ and $H = 1$ for the pole distances $\theta = 0$ degrees, 30 degrees, 60 degrees, and 90 degrees. The dotted lines indicate the voltage induced in a linear antenna placed in this field.
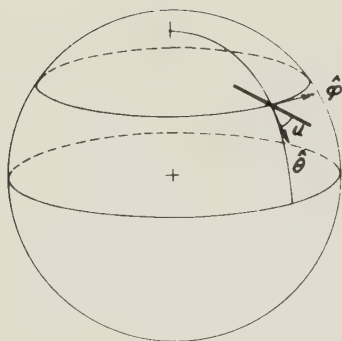
Fig. 22—Linear antenna placed in the direction (θ, ϕ) and oriented in the (θ, ϕ)-plane as indicated by the angle u.

If in the field of the ring quasi-array placed at the origin we place a linear antenna as shown in Fig. 22, the voltage induced in the linear antenna, will be given as a function of the angle u for various values of the polar distance θ by the dotted curves in Fig. 21. From

this we see that for most placings and orientations of the linear antenna the induced voltage will be different from zero. However, when the linear antenna is placed in the horizontal plane, there will be a certain direction of the antenna for which no voltage is induced in the antenna.

If the sequence of the current phase in the ring quasi-array is inverted; i.e., if we choose $H = -1$ instead of $H = 1$, as done above, the diagrams shown in Fig. 21(a)–(d) will be reflected in the θ-axis. If in any case we choose the sequence of the current phase of the ring quasi-array giving the maximum induced voltage in the linear antenna, we obtain for $\theta = 0$ degrees, 30 degrees, 60 degrees, and 90 degrees the induced voltages in the linear antenna given by the curves in Fig. 23(a)–23(d). It appears from this figure that the induced voltage varies only to a small degree with the placing and the orientation of the antenna when reversal of the current phase is applied.

a $\theta = 0°$

b $\theta = 30°$

c $\theta = 60°$

d $\theta = 90°$
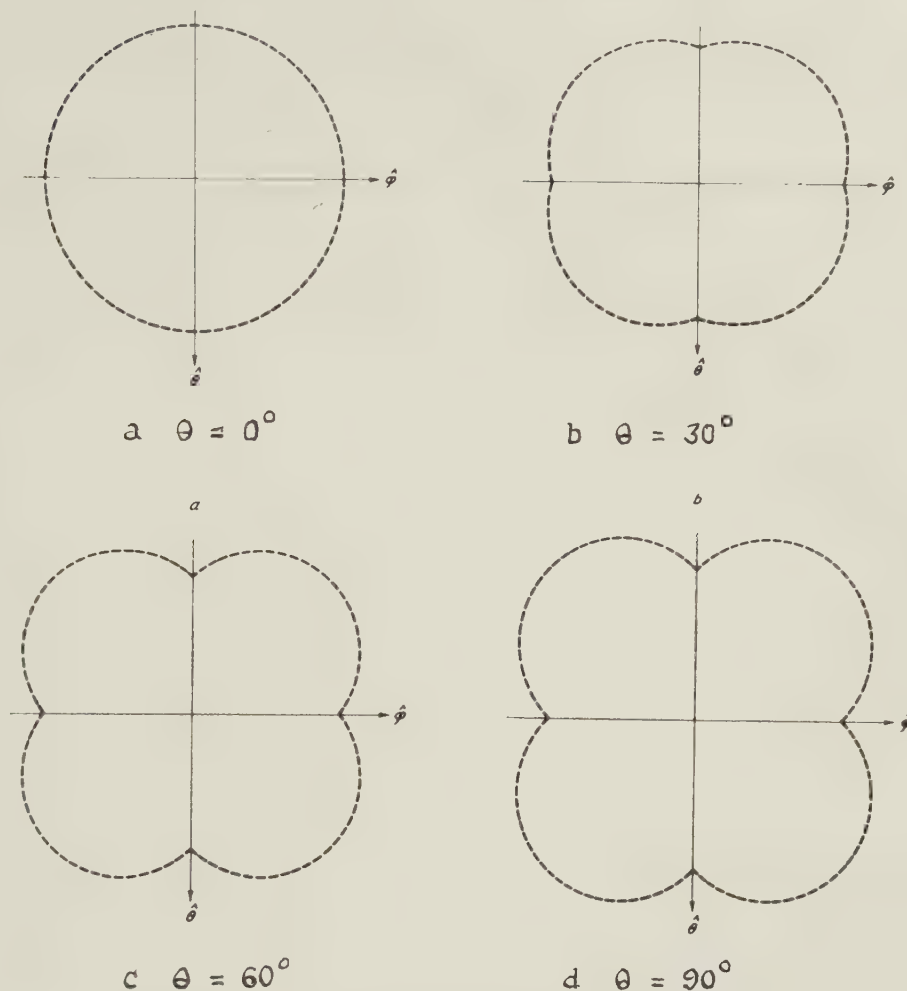
Fig. 23—Induced voltage in a linear antenna placed in the field of a homogeneous ring quasi-array of oblique dipoles with $\gamma = \pi/2$ and $H = 1$, when reversal of the sequence of the current phase is allowed.

As is well known, it is not possible to design an isotropic antenna.[28] Various attempts have been made to design antennas the properties of which approximate those of an isotropic antenna. However, for all of these antennas it is true that at least for one direction in space towards, and for one orientation of a linear receiving antenna they will induce the voltage zero in this antenna. In applications where reversal of the sequence of the current phase is permissible, the ring quasi-array described here will be a good substitute for an isotropic antenna.

As a practical design of the ring quasi-array described here it will probably be expedient to use a turnstile antenna, the four arms of which are turned upwards so that they form the angle v with the vertical direction, as shown in Fig. 24. The field from such a "quadruped antenna" will to some degree deviate from the field from the ring quasi-array considered above. By a suitable choice of the length of the arms of the antenna
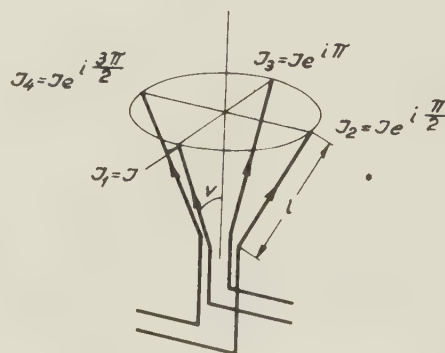
Fig. 24—Quadruped antenna.

and of the angle v it will, however, be possible to obtain a satisfactory radiation diagram.

Super-Gain

In the preceding sections it has been demonstrated that the field radiated from a homogeneous ring array or ring quasi-array will concentrate the more around

---

[28] H. F. Mathis, "A short proof that an isotropic antenna is impossible," Proc. IRE, vol. 39, p. 970; August, 1951.

the horizontal plane the larger the increment $H2\pi$ of the current phase during one revolution. By making H sufficiently large, it is possible for a homogeneous ring quasi-array of arbitrarily chosen dimensions to obtain an arbitrarily large gain. This result is inconsistent with the practical experience generally made in the antenna theory as this shows that for an antenna of given dimension it is practically impossible to obtain a gain that essentially surpasses the gain for a conventional antenna of the same diameter. Since the early forties, a phenomenon such as the one described here has aroused considerable interest among antenna theorists and has been investigated from various points of view under the name of "super-gain." As for the literature on this subject, reference is made to an earlier paper,[7] in which a critical investigation of the problem of super-gain has been made, and to another paper,[8] in which the connection between the theory of super-gain and information theory is treated.

In a few words the solution to the problem of super-gain may be thus formulated: it is certainly possible to obtain a gain that considerably surpasses the gain that is "natural" to the antenna in question; but as a result the antenna will acquire a Q so large that it will be inapplicable in practice.

For determining the approximate minimum value which a homogeneous ring array or a homogeneous ring quasi-array may have, if super-gain is not to occur, we shall base our considerations on a paper by Woodward and Lawson.[29] In this paper the authors utilize the fact that the current distribution on a linear antenna array and the array characteristic are Laplace transforms of each other. They hereby demonstrate that in order

---

[29] P. M. Woodward and J. D. Lawson, "The theoretical precision with which an arbitrary radiation-pattern may be obtained from a source of finite size," *Jour. IEE*, Pt. III, vol. 95, pp. 363–369; 1948.

to avoid super-gain for such an antenna we must demand that the greater part of the produced field is radiated at real, and not at imaginary, angles. Similarly, for the homogeneous ring array dealt with above we must demand that the greater part of the array characteristic $G \sim J_H(ka \sin \theta)$ comes within the real domain of θ i.e., for $t = \sin \theta < 1$. Consequently, we must demand that the function $J_H(kat)$ decreases rapidly towards zero for $t > 1$. In order that this may apply we must have

$$a > a_{\min.} = \frac{H\lambda}{2\pi}.$$

Accordingly, for the smallest usuable antenna array the phase of the current increases just as much per unit of length as for a conventional linear, endfire array.

By a similar argument the same result is obtained for homogeneous ring quasi-arrays of tangential and radial dipoles.

The author would like to mention the following two papers which recently came to his attention. W. Burkhardtsmaier and U. Finkbein, "Kreisgruppenantenne als Rund- und Richtstrahler," *Elektrotech.*, vol. 4, pp. 239-244, 284-290; 1950. R. H. DuHamel, "Pattern synthesis for antenna arrays on circular, elliptical, and spherical surfaces," N6-ori-71, Task XV, ONR Project No. 076161, Tech. Rep. No. 16; 1950.

Acknowledgment

Directivity, Super-Gain and Information

G. TORALDO DI FRANCIA†

Summary—In this paper some analogies between antenna theory and the theory of optical resolving power are analyzed. The effect of the finite size of a rotating antenna on the informational content of the echo is discussed, without taking into account noise. From this point of view, the most important feature of the aerial is the highest angular frequency which is contained in its radiation pattern. Super-gain is possible because no upper limit exists for this frequency. A simple method is pointed out for synthesizing a radiation pattern containing any prescribed set of finite angular frequencies. A numerical example is worked out.

INTRODUCTION

THE THEORETICAL possibility of super-gain antennas presents some paradoxical aspects. Since the discovery of super-gain, a few very good explanations of the paradox have been proposed, and one may wonder whether a new discussion of the subject is really required. However, we want to note that the apparent contradiction inherent in super-gain was due to some well-established theories of wave optics. Super-gain seemed to be at variance with the classical theory of resolving power (Rayleigh limit). This theory is most unsatisfactory even in optics, and there is nowadays a pronounced tendency towards replacing it with a more modern formulation, based on frequency analysis and information theory. It may therefore be worthwhile to analyze also the problem of super-gain from this point of view.

Antenna theory is closely related to the theory of optical diffraction patterns. This has always been realized by antenna designers who, especially in the field of microwaves, have taken advantage of the considerable experience on diffraction pattern calculations, acquired in optics for more than a century. For example in Silver's fundamental book[1] use is largely made of the aperture-field method, which is equivalent to replacing the radiating system with a suitable optical pupil. A recent and very useful review of the analogies existing between optics and antenna theory has been given by Spencer[2] who has rightly stressed the importance which such connections may have for workers in both fields. Indeed, if it is true that diffraction optics has offered a basis for many investigations on antenna theory, it is also true that optics, conversely, can benefit from the great number of new ideas that have arisen in the rapid evolution of microwave techniques. Thus, starting from the theory of super-gain antennas,

the author has been able to develop some optical super-resolving pupils.[3]

The existence of super-resolving pupils has obviously strengthened the opinion that the old theory of resolving power was defective.

It is the purpose of the present paper to point out some extremely simple facts which, though being more or less well known, can, when properly put together, clarify completely a matter which in the past has offered some mysterious and puzzling aspects. As a result, we shall be able to get a thorough understanding of the *theoretical* limitations of the performance of an aerial, as distinguished from the *practical* limitations. It will be clear that the maximum number of independent data carried by an electromagnetic echo depends on the maximum angular frequency contained in the radiation pattern of the antenna. A method will be described for obtaining a radiation pattern with any prescribed (finite) frequency.

DIRECTIONAL DEGREES OF FREEDOM AND INFORMATION

Electromagnetic radiation can be employed to collect information about the external world. In a broad sense, all systems designed to this end are optical systems. This may explain why some authors have applied information theory to optics.[4,5] An excellent discussion of its application to radar is due to Woodward.[6]

Noise is an inevitable feature of all communication systems, and is chiefly responsible for the limitation of the amount of information. However, noise depends essentially on the properties of the receiver (or receptor, in optics) and on the state of the atmosphere. In this paper we want only to stress the limitations which are inherent in the size of the aerial. Accordingly the consideration of noise will be dispensed with.

For simplicity, the treatment will be limited to a plane (say the horizontal plane). Let us first consider the case of optics. The dimension of the pupil will be designated by $2r$ (Fig. 1) and the angular semiaperture by u. The magnification will be assumed to be equal

---

† Istituto Nazionale di Ottica, Firenze, Italy.
[1] S. Silver, "Microwave Antenna Theory and Design," McGraw-Hill Book Co., *MIT.. Rad. Lab. Ser.*, New York; 1949.
[2] R. C. Spencer, "Antennas for radio astronomy," Paper presented at the Manchester Symposium on Astronomical Optics; April, 1955.
[3] G. Toraldo di Francia, "Super-gain antennas and optical resolving power," *Suppl. Nuovo Cimento*, vol. 9, pp. 426–438; 1952.
[4] A. Blanc-Lapierre, "Considérations sur la théorie de la transmission de l'information et sur son application à certains domaines de la physique," *Ann. Inst. H. Poincaré*, vol. 13, 245–296 (1953); "Upon some analogies between optics and information theory", Paper presented at the McGill Symposium on Microwave Optics, June, 1953.
[5] G. Toraldo di Francia, "Capacity of an optical channel in the presence of noise", Optica Acta, vol. 2, pp. 5–8; 1955.
[6] P. M. Woodward, "Probability and information theory with applications to radar," London, 1953.

to 1 and the instrument will be assumed to be perfect. In this condition, it is well known that the image of a point source will be represented by the complex amplitude $\sin(2\pi\alpha x/\lambda)/\pi x$, where $\alpha = \sin u$ is the numerical aperture and x the abscissa on the image plane. This diffraction pattern does not contain any spatial frequencies exceeding the absolute value α/λ. The image of an object sending out coherent light will be represented by the convolution of the complex amplitude distribution in the object and the elementary diffraction
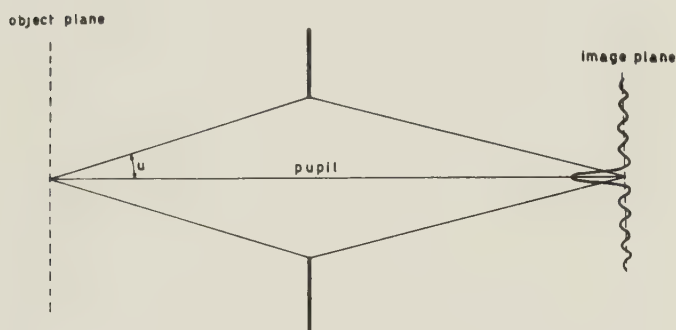
Fig. 1—Optical image information.

pattern. As a consequence, the image will not contain spatial frequencies beyond α/λ. We can now apply the sampling theorem, which represents so-to-speak the basis of information theory of continuous signals.[7] This theorem has formed the starting point of a number of applications of information theory to optical problems.[5, 8, 9] However, the author has recently emphasized that its interest may go beyond that of a starting point, inasmuch as the sampling theorem represents the only rigorous substitute for the old and unsatisfactory theory of resolving power.[10] The statement that resolving power is limited by diffraction was the first recognition that the information content of an image is limited by the finite bandwidth of spatial frequencies. But we now know that its usual formulation was wrong. As a result of the sampling theorem, the image is completely determined by its complex amplitudes at a series of points placed at a distance $\lambda/2\alpha$ apart from one another. An image having the dimension X has only $2\alpha X/\lambda$ degrees of freedom. This is the maximum number of independent data that can be derived from the image, no matter how low the noise level may be. As the object has an infinite number of degrees of freedom, while those of the image are only a finite number, we must conclude that one and the same image can correspond to many different objects. In order to select the *real* object or, at least,

in order to reduce the choice of possible objects, the observer has to rely on some past experience about the *prior* probability of a given object. The knowledge obtained from the image merely modifies this prior probability, to give a *posterior* probability. As noted by Woodward[11] it appears impossible to design a receiver capable of performing this process. However, this combination of previous information with the information supplied by the image is always more or less unconsciously operated by the observer. This is a very pertinent point, which had been completely disregarded in the old theory of resolving power. Indeed, the very expression "resolving power" has no meaning, unless the observer is assumed to have an infinite amount of prior knowledge about the object; he must know that the object consists of either *one* or *two* points. To show this, the author[10] has made use of a construction which is frequently associated with the application of the sampling theorem and which, perhaps, was first made use of by Woodward and Lawson[12] for synthesizing a given radiation pattern. Fig. 2 shows the amplitude of two equally bright points A, B, spaced $\lambda/2\alpha$ apart.
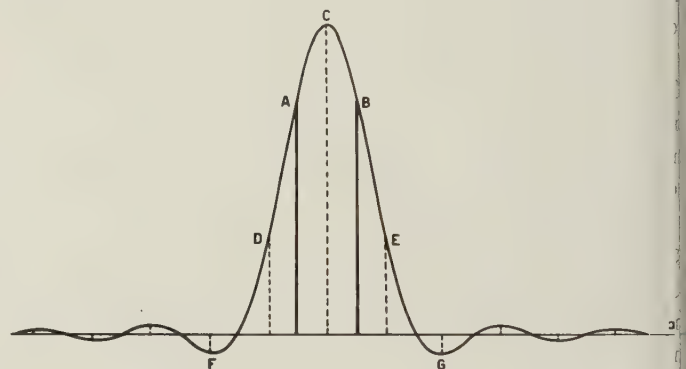
Fig. 2—Image of either two points A, B or many points C, D, E, etc.

The amplitudes of the two object points are represented by the corresponding ordinates at A and B. If the noise level is sufficiently low, the observer may be able to reveal the difference between this image and the image of a single point. If he knows that the object consists of either one point or two equally bright points, he will decide that there are actually two points A, B. If, on the contrary, he has no prior knowledge of the object, he cannot decide whether he is confronted with two points A, B or with an entire set of points C, D, E, F, etc., spaced $\lambda/2\alpha$ apart, which would produce exactly the same image. The same situation is met even in the case where one single point A is present, as is shown by Fig. 3. The observer has the choice between one point located at A or many points such as B, C, D etc. Direct

[7] C. E. Shannon, "Mathematical theory of communication," *Bell Sys.* Monograph B-1589, p. 51.

[8] D. Gabor, "Light and information," Paper presented at the Manchester Symposium on Astronomical Optics; April, 1955.

[9] E. H. Linfoot, "Noise, aberrations and the information content of optical images," Manchester Symposium; April, 1955.

[10] G. Toraldo di Francia, "Resolving power and information," *Jour. Opt. Soc. Amer.*, vol. 45, pp. 497–501; July, 1955.

[11] P. M. Woodward, *op. cit.*, p. 74.

[12] P. M. Woodward and J. D. Lawson, "The theoretical precision with which an arbitrary radiation pattern may be obtained from a source of finite size," *Journ. I.E.E.*, vol. 95, pp. 363–370; 1948.

tion finding depends essentially on the observer's knowledge that the object consists of one point only.

A similar discussion can be made in the case of incoherent illumination, as explained by the author in an earlier paper.[10]
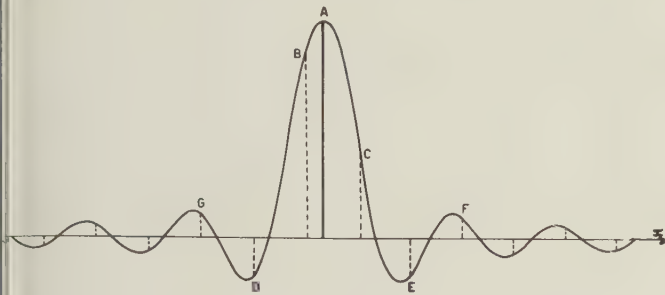
Fig. 3—Image of either one point *A* or many points *B, C, D*, etc.

In conclusion, we may say that the theory of resolving power and of direction finding is not contained in the mathematics of diffraction and can only be based on the combination of prior knowledge with the additional information provided by the image. The only limit that mathematics does set, concerns the number of degrees of freedom of the image, and not the way in which the degrees of freedom can be utilized.

It would seem evident that the foregoing considerations can be applied with only slight changes to the azimuthal resolving power of a *PPI* image obtained with a rotating antenna. This is not exactly true, as we shall see later. However we want first to outline a rough argument which might be applied in this case.

One could assume a linear aperture and start from the remark that the radiation pattern, considered as a function of the sine of the azimuth, is the Fourier transform of the amplitude distribution on the aperture. Incidentally, we note that this trivial fact seems to have been independently discovered and rediscovered by many workers in different countries, and there would be no point in establishing priorities. The author, though claiming no priority, may mention that he made extensive use of this property in a series of articles,[13] which, apparently, failed to be noticed at the time of their publication. A good review of Fourier techniques applied to radiation patterns has been given by Spencer.[14] Because of the reciprocity of Fourier transforms, the aperture field is in turn the transform of the radiation pattern, and since the aperture has a finite size, the radiation pattern is frequency limited. Now, as the aerial turns, the echo from the object (when the latter is supposed to be located on a circle with the center on the aerial) is represented by the convolution of the power spectrum of the aerial and the reflection co-

efficient of the object. As a consequence, the spectrum of the echo will also be frequency limited, and by applying the sampling theorem, we find that the echo is completely determined by its values at a discreet set of angles.

From this argument it would follow that the echo can never contain more than a given finite number of independent data. A knowledge of the noise level can tell us how much of this information content is utilized and how much is lost. However, the only way to increase the number of independent data available would be to increase the size of the antenna.

Now, the foregoing argument is not correct, because the convolution variable should be an angle, while the variable of the Fourier transform is the sine of an angle. Indeed, both the linear aperture and the sine variable, which are most suitable for the treatment of many optical problems, give rise to some difficulties in the problem of the rotating antenna. As we shall see later, it is more natural to consider a circular aperture and to take the azimuth as variable.

Super-Gain and Angular Frequencies

The first evidence that directivity is not completely limited by the size of the antenna was due to Hansen and Woodyard[15] and to Schelkunoff.[16] A little later Bouwkamp and de Bruijn[17] proved that there is no theoretical upper limit to the gain obtainable from a linear antenna. Dolph[18] and Riblet[19] suggested a mathematical method for designing arrays with prescribed beamwidths and sufficiently low side-lobe level. These results have been since extended and applied by many workers. Recent contributions are due to Duhamel,[20] Rhodes,[21] Stegen,[22] and van der Maas.[23] An interesting case of super-gain for ring quasi-arrays has been pointed out by Knudsen.[24] The author has considered the case of a few thin and concentric rings[3] and the case of an entire circular aperture divided into a number of circular zones.[25]

[13] G. Toraldo di Francia," "Saggio su una teoria generale dei reticoli," *Ottica*, vol. 6, pp. 287–307; 1941; "Alcuni fenomeni di diffrazione trattati mediante il principio dell'interferenza inversa," *Ottica*, vol. 7, pp. 117–136; 1941.

[14] R. C. Spencer, "Some Useful Operational and Fourier Techniques," paper presented at the McGill Symposium on Microwave Optics; June, 1953.

[15] W. W. Hansen and J. R. Woodyard, "A new principle in directional antenna design," *Proc. IRE*, vol. 26, pp. 333–345; 1938.

[16] S. A. Schelkunoff," A mathematical theory of linear arrays," *Bell Sys. Tech. Jour.*, vol. 22, pp. 80–107; 1943.

[17] C. J. Bouwkamp and N. G. de Bruijn, "The Problem of Optimum Antenna Current Distribution," *Philips Res. Rep.*, vol. 1, pp. 135–158; 1946.

[18] C. L. Dolph, "A current distribution for broadside arrays, which optimizes the relationship between beamwidth and side-lobe level," *Proc. IRE*, vol. 34, pp. 335–348; June, 1946.

[19] H. J. Riblet, "Discussion on Dolph's paper," *Proc. IRE*, vol. 35, pp. 489–492; May, 1947.

[20] R. H. Duhamel, "Optimum patterns for endfire arrays," *Proc. IRE*, vol. 41, pp. 652–659; May, 1953.

[21] D. R. Rhodes, "The optimum linear array for a single main beam," *Proc. IRE*, vol. 41, pp. 793–794, June, 1953.

[22] R. J. Stegen, "Excitation coefficients and beamwidths of Tchebycheff arrays," *Proc. IRE*, 41, 1671–1679; November, 1953.

[23] G. J. van der Maas, "A simplified calculation for Dolph-Tchebytcheff arrays," McGill Symposium on Microwave Optics; June, 1953.

[24] H. L. Knudsen, "The field radiated by a ring quasi-array of an infinite number of tangential or radial dipoles," *Proc. IRE*, 41, pp. 781–789; June, 1953; "Radiation resistance and gain of homogeneous ring quasi-array," *Proc. IRE*, 42, 686–695; April, 1954.

[25] G. Toraldo di Francia, "Nuove pupille superrisolventi," *Atti Fond. G. Ronchi*, vol. 7, pp. 366–372; 1952.

A common defect of all these antennas is that they require enormous reactive currents and have very high Q. In most cases they cannot be realized.

Anyhow, apart from the question of practical realization, it is certain that there are a number of different ways to design an antenna of any desired beamwidth with any desired side-lobe level. Since this seems to be at variance with previous theoretical results, it is interesting to investigate what was wrong in those results.

The classical statement that maximum gain is given by a uniform field distribution over the aperture[26] was wrong simply because its proof rested on the assumption that the total radiated power is equal to the integral of the squared amplitude over the pupil. Indeed, such an integral includes both the power actually radiated by the aperture and the reactive power corresponding to evanescent waves. Woodward and Lawson[12] were the first to call attention to the role of evanescent waves in super gain problems, while the author[27] was probably the first to call attention to their importance in all diffraction phenomena.

The discussion of the foregoing section points out that another question must be answered before we get a clear picture of the informational limitations inherent in an antenna of given size. We must ask ourselves whether or not the angular frequencies of the echo really have a strict upper limit. It would seem on first sight that any desired frequency should be obtainable by means of a super-gain antenna. That this is actually true can be most readily proved by considering a circular aperture in place of the straight aperture and taking the azimuth as variable in place of its sine.

Let us consider a circular aperture of radius r in the horizontal plane (Fig. 4). We want to evaluate the beam amplitude $A(\theta)$ in a direction CD corresponding to the azimuth θ. Somewhat similar problems have been treated by Carter,[28] Walsh,[29] Knudsen,[24] and others. If $a(\theta)$ designates the complex amplitude over the aperture, its value at point P will be given by $a(\theta - \pi/2 + \varphi)$. The retardation corresponding to P is readily found to be $-r \sin \varphi$, so that we obtain

$$A(\theta) = \int_0^\pi a(\theta - \pi/2 + \varphi) \exp(-ikr \sin \varphi) f(\varphi)\, d\varphi, \quad (1)$$

where $f(\varphi)$ is a function representing the directivity of the aperture elements. This factor depends on the actual constitution of the antenna. We now note that $a(\theta)$ is necessarily a periodic function, which can be expressed by the Fourier series

[26] R. K. Luneberg, "Mathematical theory of optics," Providence, p. 391; 1944.

[27] G. Toraldo di Francia, "Le onde evanescenti nella diffrazione," *Ottica*, vol. 7, pp. 197–206; 1942; M. Schaffner and G. Toraldo di Francia, "Microoonde evanescenti generate per diffrazione," *Nuovo Cimento*, vol. 6, pp. 125–131; 1949.

[28] P. S. Carter, "Antenna arrays around cylinders," PROC. IRE, 31, pp. 671–693; 1943.

[29] J. E. Walsh, "Radiation pattern of arrays on a reflecting cylinder," PROC. IRE, vol. 39, pp. 1074–1081; September, 1951.

Fig. 4—Radiation from a circular aperture.

$$a(\theta) = \sum_{-\infty}^{+\infty} a_n \exp(in\theta) \quad (2)$$

Upon substitution of (2) into (1), we obtain;

$$A(\theta) = \sum_{-\infty}^{+\infty} \exp(in\theta)(-i)^n a_n \int_0^\pi \exp[i(n\varphi - kr \sin \varphi)]f(\varphi)\, d\varphi \quad (3)$$

Putting:

$$A_n = (-i)^n a_n \int_0^\pi \exp[i(n\varphi - kr \sin \varphi)]f(\varphi)\, d\varphi \quad (4)$$

we may also write;

$$A(\theta) = \sum_{-\infty}^{+\infty} A_n \exp(in\theta) \quad (5)$$

This equation expresses $A(\theta)$ as a Fourier series. The coefficients A_n are given by (4) and are each proportional to the corresponding a_n.

To go further, we must specify $f(\varphi)$. We shall assume $f(\varphi)$ to have the general form $\sin^p \varphi$ with p integral and positive. The case $p = 0$ could be approximated by an array of vertical dipoles on a cylindrical absorber and the case $p = 1$ by an array of vertical dipoles on a reflecting cylinder. For $p = 0$, the integral in (4) is of a well-known type and is expressed by:[30]

$$\frac{1}{\pi} \int_0^\pi \exp[i(n\varphi - kr \sin \varphi)]\, d\varphi = J_n(kr) + iE_n(kr) \quad (6)$$

where $J_n(kr)$ represents Anger's function and $E_n(kr)$ Weber's function of order n. In our case, n is an integer and Anger's function coincides with the Bessel function of the same order. By differentiating (6) p times with respect to kr, we obtain

$$\frac{(-i)^p}{\pi} \int_0^\pi \exp[i(n\varphi - kr \sin \varphi)] \sin^p \varphi\, d\varphi$$
$$= J_n^{(p)}(kr) + iE_n^{(p)}(kr). \quad (7)$$

[30] G. N. Watson, "Theory of Bessel functions," Cambridge, p. 312; 1922.

In conclusion, from (4) and (7) we get

$$a_n = C_n A_n \qquad (8)$$

with

$$C_n = \frac{i^{n-p}}{\pi[J_n^{(p)}(kr) + iE_n^{(p)}(kr)]}. \qquad (9)$$

Eq. (8) and (9) enable us to evaluate the coefficients a_n of the aperture illumination when the coefficients A_n of the radiation pattern are prescribed.

As a result, we may state that there is no upper limit for the angular frequencies of the radiation pattern. The radiation pattern can contain any arbitrary finite frequency, consequently there is no theoretical limit for the number of sampling points which are necessary to specify the echo from the object.

However, we must make an important remark. The following series expansions of J_n and E_n are well known:

$$J_n(x) = \frac{(x/2)^n}{0! \, n!} - \frac{(x/2)^{n+2}}{1! \, (n+1)!} + \frac{(x/2)^{n+4}}{2! \, (n+2)!} - \cdots . \qquad (10)$$

$$E_n(x) = \frac{1 - \cos n\pi}{n\pi} \left[1 - \frac{x^2}{2^2 - n^2} \right.$$
$$\left. + \frac{x^4}{(2^2 - n^2)(4^2 - n^2)} - \cdots . \right]$$
$$- \frac{1 + \cos n\pi}{\pi} \left[\frac{x}{1^2 - n^2} - \frac{x^3}{(1^2 - n^2)(3^2 - n^2)} + \cdots \right]. \qquad (11)$$

When these expansions are substituted into (9) for the case $p = 0$, it is seen that for a fixed value of kr and increasing values of n, the absolute values of the coefficients C_n become greater than a quantity proportional to n. This means that very high frequencies in the radiation pattern require very large currents of the same angular frequency in the aerial. In particular, there follows also that if $A(\theta)$ were completely prescribed in advance, the Fourier series obtained for $a(\theta)$ would, generally speaking, be divergent. The series must therefore be cut off at a prescribed frequency. These results are similar to those obtained by Chu,[31] for the gain of an omnidirectional antenna in the meridional plane and by Leonard[32] for the synthesis of the aperture distribution on an infinite slot.

The situation is even worse for $p \neq 0$. By differentiating (10) and (11) p times and substituting into (9), it is readily seen that for increasing values of n, the coefficients C_n tend to become greater than a quantity proportional to n^{p+1}.

The foregoing results will be illustrated by a numerical example in the case $p = 0$. An ideally directive pattern would be represented by a δ function of θ. Accordingly, we shall require that $A(\theta)$ be a normalized δ function of the form

[31] L. J. Chu, "Physical limitations of omnidirectional antennas," *Jour. Appl. Physics.*, vol. 19, pp. 1163–1175; 1948.

[32] D. J. Leonard, "The synthesis of the aperture field from an arbitrary radiation pattern," Paper presented at the McGill Symposium in Microwave Optics; June, 1953.

$$A(\theta) = \frac{1}{2N + 1} \sum_{-N}^{+N} \exp(in\theta). \qquad (12)$$

In the limit, for $N = \infty$ this expression has the value 1 for $\theta = 0$ and vanishes elsewhere. Here, by comparison with (5), we have $A_n = 1/(2N + 1)$ for $|n| \leqslant N$ and $A_n = 0$ for $|n| > N$. From (2) and (8) there follows

$$a(\theta) = \frac{1}{2N + 1} \sum_{-N}^{+N} C_n \exp(in\theta).$$

Now, from (9), (8) and (11) one can readily verify that $C_{-n} = C_n$, so that

$$a(\theta) = \frac{1}{2N + 1} (C_o + 2C_1 \cos \theta + 2C_2 \cos 2\theta + \cdots + 2C_N \cos N\theta). \qquad (13)$$

Unfortunately, to the author's knowledge, no extensive tables of Weber's functions for $n > 1$ are available. However, the series (11) is rapidly convergent when the argument is not too large, and can be used for the calculation.

The value of kr has been taken for unity. This corresponds to an aerial of diameter λ/π which, according to classical theories, should have an extremely poor directivity.

Four successive approximations to the δ function radiation pattern are shown in Fig. 5; they have been calculated by means of (12) for $N = 1, 2, 3, 4$ respectively. The corresponding aperture distributions have been calculated with (13) and are shown in Fig. 6. The amplitude is represented by the solid line. For reasons of continuity, the amplitude has been allowed to take both positive and negative values; a passage from positive to negative values or vice-versa means a reversal of phase. The phase is represented by the dashed line and does not take into account the said reversals; besides, an arbitrary constant has been added to the phase in each case, so as to start with zero phase at $\theta = 0$.

The amplitude shows a number of oscillations and the phase a number of steps. Both numbers increase with increasing N. It is to be noted that the order of magnitude of the amplitude does not change very much up to $N = 3$. Only with $N = 4$, the increase of the amplitude becomes considerable (note the change of scale in the figure).

Conclusion

In this paper, we have pointed out the inadequacy of resolving power to represent the informational limitation inherent in the finite size of an antenna or of an optical pupil. It is impossible to speak of resolving power, without taking into account both the noise level in the receiver and the amount of prior information about the object, which is possessed by the observer.

The only limitation which is really found by mathematics, concerns the maximum number of independent data which are necessary to specify a *PPI* image or the echo from a circular object. This number, in turn, de-
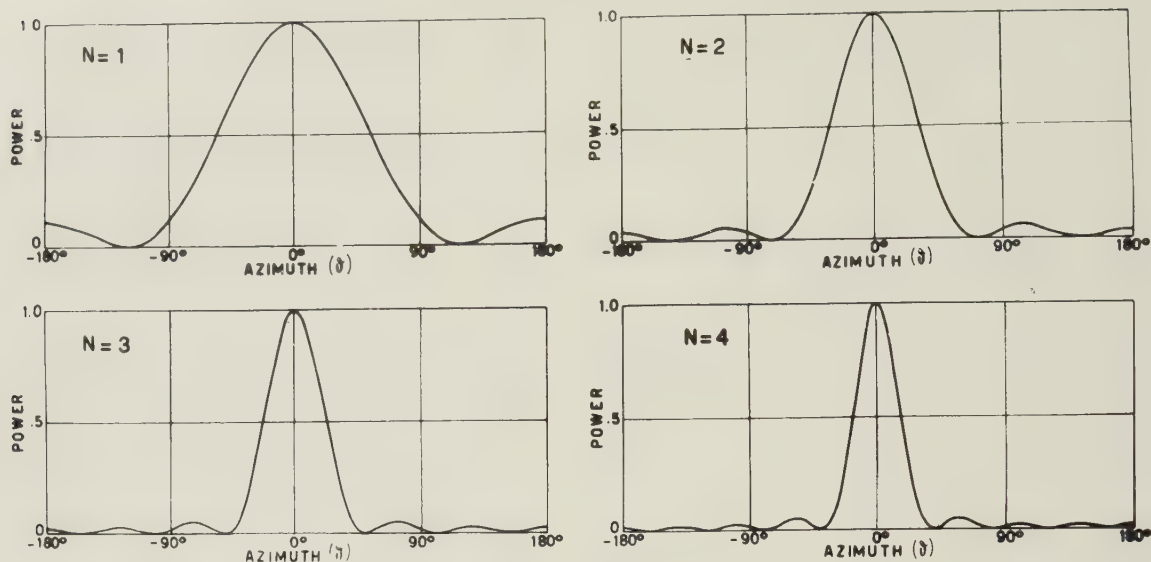
Fig. 5—Four successive approximations to an ideally directive radiation pattern.
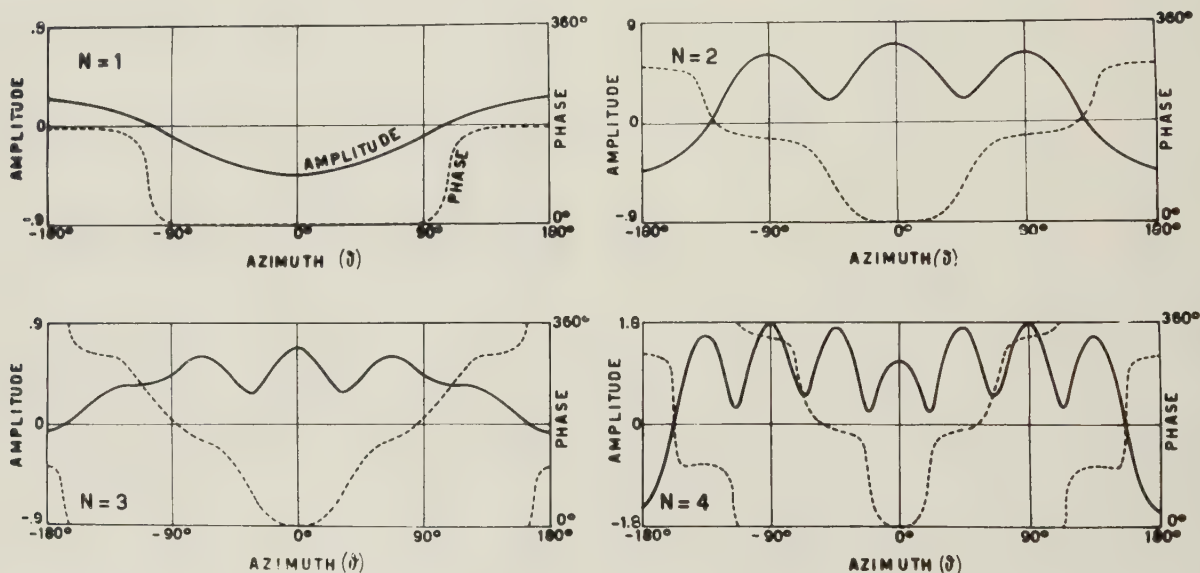
Fig. 6—Amplitude and phase distributions on the circular aperture for obtaining the corresponding radiation patterns of Fig. 5.

pends on the maximum angular frequency which is contained in the radiation pattern.

Angular frequencies are not too obviously related to the sine variable which is ordinarily utilized in the case of a linear aperture. Linear apertures and Fourier transforms which, as it were, represent the natural way to treat optical problems, appear less natural in the case of a rotating antenna.

If the straight aperture is replaced by a circular

aperture and the sine variable by an angular variable, the problem of super gain becomes a trivial matter. It is indeed very easy to synthesize an aperture distribution giving rise to any desired angular frequency in the radiation pattern.

In conclusion, the finite size of the antenna cannot by itself set any strict limitation on the information content of the echo. Such a limitation can be found solely by taking into account the influence of noise.

Exact Treatment of Antenna Current Wave Reflection at the End of a Tube-Shaped Cylindrical Antenna

ERIK HALLÉN†

INTRODUCTION

ANTENNAS, whose radiating body has a cross section which is small in comparison to the length of the antenna and to the wave length; i.e., antennas consisting of wires and rods, are to a very wide extent treated with the help of a linearized integral equation, the invention of the author.[1-3] In this equation the distance between two points on the antenna is normally represented by the distance between the corresponding points on some central line. Only when the distance is small this is not permitted and from such regions arises the only term which contains the dimension of the cross section, which is a parameter mainly consisting of a logarithm. The equation therefore has a certain limited degree of accuracy which is such that the ratio of the radius of cross section to the length of the antenna or to the wavelength is neglected compared with unity. The results which can be drawn from the linearized integral equation thus also should have this limited accuracy which is a normal one in electrotechnics in all kinds of devices, where wires are involved. Nevertheless much discussion has gone on about this accuracy. The only way of finding definite numerical answers to this question is to solve exactly the antenna integral equations, both the linearized one and the exact one, for some special case. Nowadays this can be done for a straight cylindrical tube-shaped antenna.

(R. Gans has recently expressed the opinion that Hallén's linearized integral equation should not have any *exact* solution.[4,5] This is a mistake made by Gans because he apparently has never seen my original papers. What he studies and criticizes is the coarser form of the equation given in many American papers and books as "Hallén's integral equation." Gans in reality criticizes the deviation that is made in those papers from my own form, which is not subject to any criticism of the kind expressed by Gans.[6,7]

It must also be remembered that even if I would agree with Gans, to some extent, if he had directed his criticism to the proper persons, it is for practical purposes not very essential if the coarser form of the equation is used in the beginning. Those authors have at a later stage in their deductions usually completed the linearization so that they in reality use the correct form even if it has not been written explicitly. As they try to find only an approximate solution it is even for this reason unessential if their equation is not quite the correct linearized one.)

Any electric field in free space, periodic with respect to time as $e^{j\omega t}$, can be expressed with the help of a vector potential A as follows:

$$E = \frac{c}{j\beta} (\text{grad div } A + \beta^2 A), \qquad (1)$$

where c is the electromagnetic wave velocity and

$$\beta = \frac{\omega}{c}, \qquad (2)$$

a wave constant. On a cylindrical antenna oscillating under conditions which involve axial symmetry, the current and consequently its vector potential has the direction of the cylinder axis. If the ohmic resistance is neglected the tangential component of E vanishes in the surface of the antenna and from (1) follows the nowadays well-known fact[1-3] that the vector potential, and hence the scalar potential, of the antenna field is exactly sine-shaped along the surface of a cylindrical antenna. This gives the integral equation for the antenna current $I(x)$ of a cylindrical antenna, fed by a potential jump $2V_o$ in the middle (reference 3, eq. 35.15):

$$\int_{-l}^{l} d\xi I(\xi) \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi$$
$$= \frac{4\pi}{Z_o} (-V_o e^{-i\beta|x|} + A \cos \beta x), \qquad (3)$$

where $2l$ is the length, a the cylinder-radius of the antenna,

$$r = [(x - \xi)^2 + 4a^2 \sin^2 \tfrac{1}{2}\varphi]^{1/2},$$

the distance between two surface points, Z_o the wave resistance of free space ($Z_o = 377$ ohms), and A an unknown constant. The minus sign on V_o, which we have introduced here, although it was not in the previous papers, only indicates that we have turned the potential jump so that the outgoing potential wave is positive when traveling in the direction of negative x. In order that (3) shall be mathematically exact the ohmic resistance should be negligible and the cylinder should have no end surfaces; i.e., it must be a thin tube.

† Royal Institute of Technology, Stockholm, Sweden.
[1] E. Hallén, "Über die Elektrischen Schwingungen in Drahtförmigen Leitern," Uppsala Universitets Arsskrift, pp. 1–102; 1930.
[2] E. Hallén, "Theoretical investigations into the transmitting and receiving qualities of antennae"—N. A. R. Soc. Sci. Upsal., series IV, vol. 11, pp. 1–44; 1938.
[3] E. Hallén, "Elektricitetslära," Stockholm, pp. 1–556; 1953. (In Swedish.)
[4] R. Gans, "Zur theorie der geradlinigen antenne," Arch. Elektr. Übertragung, vol. 7, pp. 169–180; 1953.
[5] R. Gans and M. Bemporad, "Contribution to the theory of the rectilinear antenna," "New Research Techniques in Physics," Rio de Janeiro; 1954.
[6] C. H. Papas, Math. Rev. vol. 15, pp. 1007–1008; 1954.
[7] F. E. Borgnis, and C. H. Papas, "Randwertprobleme der Mikrowellenphysik," Springer Verlag, p. 225; 1955.

(In practice only the pieces nearest to the cylinder ends need to be hollow.)

The current $I(x)$ can be considered either as a standing wave or as a system of traveling waves, one outgoing current wave traveling in both directions from the feeding point, and a series of current waves reflected at the ends of the antenna. The terms of the right side of (3) are, besides a constant factor, the corresponding terms of the vector potential along the antenna. Thus $-(V_o/c)e^{-i\beta|x|}$ is the outgoing vector potential wave and $A/c \cos \beta x$ the sum of all the reflected vector potential waves. The outgoing scalar potential wave is $+V_o e^{i\beta x}$ in the direction of negative x and $-V_o e^{-i\beta x}$ in the positive direction. The sum of all the reflected scalar potential waves is $-jA \sin \beta x$. Thus the traveling potential waves have constant amplitudes (on the surface of the cylindrical antenna), whereas the corresponding traveling current waves always decrease.

The outgoing traveling wave is easily found from (3). It is the same as the current on an infinite antenna without end reflections, hence $A = 0$ and $l = \infty$. Thus the outgoing current wave satisfies the integral equation[3,8].

$$\int_{-\infty}^{\infty} d\xi I(\xi) \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi = -\frac{4\pi}{Z_o} V_o e^{-i\beta|x|}. \quad (4)$$

[For the solution of this equation see (9).] Before we proceed we will even write down the linearized equations, which correspond to (3) and (4). These are for an antenna of finite length[2] (Eq. 24):

$$I(x) \log \frac{4(l^2 - x^2)}{a^2} + \int_{-l}^{l} \frac{I(\xi)e^{-i\beta|x-\xi|} - I(x)}{|x - \xi|} d\xi$$

$$= \frac{4\pi}{Z_o} (-V_o e^{-i\beta|x|} + A \cos \beta x), \quad (5a)$$

or in a modified form[3]:

$$I(x)\left[\log \frac{4}{a^2\beta^2} - 2\gamma - j\pi - e^{-i\beta(l-x)}\mathbf{l}(\beta(l - x))\right.$$

$$\left. -e^{-i\beta(l+x)}\mathbf{l}(\beta(l + x))\right] + \int_{-l}^{l} \frac{I(\xi) - I(x)}{|x - \xi|} e^{-i\beta|x-\xi|} d\xi$$

$$= \frac{4\pi}{Z_o}(-V_o e^{-i\beta|x|} + A \cos \beta x), \quad (5b)$$

where the function $\mathbf{l}$ is the complex amplitude function of the sine and cosine integral as defined (and tabulated) earlier by the author.[8–10] From (5b) we immediately get the linearized integral equation for the outgoing waves, if we put $l = \infty$ and $A = 0$.[3,8]

[8] E. Hallén, "Properties of a long antenna," *Jour. Appl. Phys.*, vol. 19, pp. 1140–1147; 1948.

[9] E. Hallén, "Travelling Waves and Unsymmetrically Fed Antennas," Tech. Rep. 49, Cruft Lab., Harvard University, Cambridge, Mass., pp. 1–48; 1948.

[10] E. Hallén, "Further investigations into iterated sine and cosine integrals and their amplitude functions with reference to antenna theory" *Trans. Roy. Inst. Tech.*, Stockholm, no. 89, pp. 1–43; 1955.

$$I(x)\left[\log \frac{4}{a^2\beta^2} - 2\gamma - j\pi\right]$$

$$+\int_{-\infty}^{\infty} \frac{I(\xi) - I(x)}{|x - \xi|} e^{-i\beta|x-\xi|} d\xi = -\frac{4\pi}{Z_o} V_o e^{-i\beta|x|} \quad (6)$$

The difference between (6) and (4) is that in (6) the radius of the cross section of the antenna is consequently suppressed except in the first logarithmic term. There is the same connection between (5a), (5b) and (3). The linearized equations are much easier to handle, but that is not the main reason to use them as we can now even solve (3) and (6) exactly. The linearizing has been invented in order to make it possible to solve antenna problems of much more complicated nature. Even if the antenna has a curved instead of a straight central line, has variable cross section and cross section which is not circular, integral equations can still be set up with only one unknown variable of a type corresponding to (5a). The fact that in the general case the vector potential along the antenna is not sine-shaped, presents no obstacle. If even the ohmic resistance of the antenna is taken into account as well as an incoming outer field, this general linearized antenna equation can be written:[1,2]

$$I(x) = \frac{1}{\Omega} \left\{ -I(x) \log \frac{l^2 - x^2}{l^2} \right.$$

$$-\int_{-l}^{l} \left(\frac{I(\xi)e^{-i\beta r(\xi, x)}}{r(\xi, x)} - \frac{I(x)}{|x - \xi|}\right)d\xi$$

$$+\int_{-l}^{l} I(\xi) d\xi \int^{x} \left[\beta \sin \beta(x - s) \frac{1 - \cos (\xi, s)}{r(\xi, s)} e^{-i\beta r(\xi,s)}\right.$$

$$\left. + \cos \beta(x - s) \left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial s}\right) \frac{e^{-i\beta r(\xi,s)}}{r(\xi, s)}\right] ds$$

$$- 2\int^{x} I(s) \cos \beta(x - s) \frac{a'(s)}{a(s)} ds$$

$$+ j\frac{4\pi}{Z_o} \int^{x} I(s) \sin \beta(x - s)z_\beta(s) ds$$

$$-j\frac{4\pi}{Z_o} \int^{x} E_o(s) \sin \beta(x - s) ds + j\frac{4\pi}{Z_o} V_o \sin \beta |x - l_1|$$

$$+ A \cos \beta x + \overset{.}{B} \sin \beta x \quad (5c)$$

where $\Omega = 2 \log 2l/a(x)$ is a parameter. (5d)

Here x and ξ are the lengths of the arc of the central line of the antenna reaching from $-l$ to l; $r(\xi, x)$ is the straight distance between two points on the central line (thus $r(x, x) = 0$); $a(x)$ is the radius of cross section at the point x, if circular, and the "equivalent radius" if not; z_β is the inner impedance per unit length of the antenna wire. $E_o(x)$ is the component along the antenna central line of an incoming outer electric field if any, and $2V_o$ the driving potential jump [directed as in (3)], if any, applied in the point $x = l_1$; and A and constants. The integrals with respect to s, which have upper limits x, have arbitrary, but constant lower limits. All integrands remain finite even when $\xi = $ or $\xi = s$.

Eqs. (4) and (6) for the outgoing current wave both can easily be exactly solved. The general solution of both includes an undetermined solution to the corresponding homogeneous equations; i.e., the equations with the right side put equal to zero. In the case of the exact equation (4) this eigen-solution (see further below) is the wave guide solution for the tube:

$$I = \text{constant} \cdot e^{\pm \frac{x}{a} \sqrt{\xi_m{}^2 - a^2 \beta^2}}, \qquad (7)$$

(reference 3, page 410) where the ξ_m are the zeros of the Bessel function of order 0. This solution in all ordinary antenna cases ($\lambda \gg a$), when the wavelength is big enough to make $\beta < \xi_m/a$, is aperiodic in space: frequency below cutoff frequency of the tube. In the case of the linearized antenna integral equation (6) we have (see further below) a corresponding eigen-solution to the homogeneous equation:

$$I = \text{constant} \cdot e^{\pm i \frac{x}{a} \sqrt{\frac{4}{\gamma_1{}^2} + a^2 \beta^2}}, \qquad (8)$$

where $\gamma_1 = 1.781072$ or $\log \gamma_1 = \gamma = 0.577216$ is Euler's constant. Eq. (8) represents traveling waves in both directions, which in all ordinary antenna cases ($\lambda \gg a$) would have a very low velocity. These waves have no physical significance.

Thus the general solutions of both (4) and (6) are undetermined by terms of the kind indicated by (7) and (8). However, if we add the condition that the solutions to (4) and (6) should have symmetry with respect to the feeding point, so that the two waves, going out in both directions, should be equivalent, then the solution of (4) as well as of (6) is unique. This solution of the exact equation (4) is (see reference 3, eq. 35.24; reference 8, eq. 8):

$$I(x) = -\frac{4\pi}{Z_o} V_o \frac{j\beta}{2\pi}$$

$$\cdot \int_\Gamma \frac{e^{i\alpha x}}{(\alpha^2 - \beta^2) I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha, \qquad (9)$$

and of the linearized equation (6),

$$I(x) = -\frac{4\pi}{Z_o} V_o \frac{j2\beta}{2\pi} \int_\Gamma \frac{e^{i\alpha x}}{(\alpha^2 - \beta^2) \log \frac{4}{a^2 \gamma_1{}^2 (\alpha^2 - \beta^2)}} d\alpha, \qquad (10)$$

where I_o and K_o[11] (pp. 78–79) are the Bessel and Hankel functions of imaginary argument and Γ an integration path in the complex α-plane which follows the real axis from $-\infty$ to $+\infty$ but avoids $\alpha = \pm\beta$ as if β had a small negative imaginary part. If x is changed into $-x$, (9) and (10) remain unaltered.

Both (9) and (10) are exact solutions respectively of (4) and (6), but they still do not represent exactly pure antenna waves. Both contain small terms of a type similar to (7) and (8) respectively but with the difference that x is replaced by $|x|$; i.e., (9) contains

outgoing waveguide waves (below cutoff frequency) and (10) an extra outgoing very slowly traveling wave. Those additional parts in (9) and (10) come from the poles of the integrand. By changing the path of integration we can separate those small terms from the rest, which is the true antenna outgoing traveling wave, for (9), according to[3] (Reference 3, eq. 35. 28a):

$$I(x) = -\frac{4\pi}{Z_o} V_o \psi(\beta x) e^{-i\beta |x|}, \qquad (11)$$

where

$$\psi(\beta x) = e^{i\beta |x|} \frac{2}{\pi^2}$$

$$\cdot \left\{ \int_0^1 \frac{e^{-i\beta |x| \sqrt{1 - u^2}}}{u\sqrt{1 - u^2} H_o{}^{(1)}(a\beta u) H_o{}^{(2)}(a\beta u)} du \right.$$

$$\left. + j \int_1^\infty \frac{e^{-\beta |x| \sqrt{u^2 - 1}}}{u\sqrt{u^2 - 1} H_o{}^{(1)}(a\beta u) H_o{}^{(2)}(a\beta u)} du \right\}, \qquad (12)$$

and in the case (10), (reference 3, eq. 35.36):

$$\psi(\beta x) = e^{i\beta |x|} \frac{2}{\pi^2} \left\{ \int_0^1 \frac{e^{-i\beta |x| \sqrt{1 - u^2}}}{u\sqrt{1 - u^2} [1 + \frac{4}{\pi^2} \log^2(\frac{1}{2} a\beta\gamma_1 u)]} du \right.$$

$$\left. + j \int_1^\infty \frac{e^{-\beta |x| \sqrt{u^2 - 1}}}{u\sqrt{u^2 - 1} [1 + \frac{4}{\pi^2} \log^2(\frac{1}{2} a\beta\gamma_1 u)]} du \right\}. \qquad (13)$$

(I repeat that the minus signs before V_o in (3), (4), (5), (6), (9), (10), and (11) only indicate that whereas the current is counted positive in the direction of positive x, the potential jump $2V_o$ is directed so that it would drive current in the negative direction; i.e., the scalar potential wave is positive for negative x and vice versa.) Another way of separating the true antenna current from the total current follows from (17c) and (17d). In both cases the function $\psi(\beta x)$, in spite of the first factor is perfectly aperiodic. (This has been diagrammed in reference 3, p. 415). In the case of the exact solution (9), and (11, 12) have each a distinct physical meaning: (11, 12) is the true antenna current on the outside of the antenna tube, whereas (9) is the sum of this current and the waveguide current inside the tube. In the linearized case there is no such distinction. The new expressions (11, 12), and (13) still are exact solutions of the integral equations (4) and (6) but only on one side of the feeding point, either the side of positive x or of negative x; for the other side it is necessary to add an eigen-solution of the type (7) respective (8) to make the integral equation satisfied for all values of x. The expressions (9) and (10) are the only symmetrical solutions to (4) and (6). We remark that the terms which make the difference between (9) and (11, 12) and between (10) and (11, 13) contain the small factor $a\beta$.

The close connection between (12) and (13) or between (9) and (10) is obvious. We have in the de-

---

[11] G. N. Watson, "A Treatise on the Theory of Bessel Functions," 2nd Ed., Cambridge University Press, Cambridge, pp. 1–804; 1944.

nominator of the integrands in (12)

$$H_o^{(1)}(z)H_o^{(2)}(z) = J_o^2(z) + Y_o^2(z)$$

$$= 1 + \frac{4}{\pi^2}\log^2 \tfrac{1}{2}\gamma_1 z + z^2\left\{-\tfrac{1}{2} + \frac{2}{\pi^2}\left(\log \tfrac{1}{2}\gamma_1 z\right.\right.$$
$$\left.\left. - \log^2 \tfrac{1}{2}\gamma_1 z\right)\right\} + \cdots,$$

and thus we can get (13) from (12) by suppressing the powers of $a\beta$ but keeping it in the logarithm. This is just what we did when we put up the linearized integral equation (5a). The numerical agreement between the solution (12) of the exact equation (4) and the solution (13) of the linearized equation (6) is very high even near the feeding point, in all ordinary cases. Only for extremely thick antennas is there a difference. It now also appears that the consequent complete linearization of the integral equation (5a) with suppressing of all powers of a (but keeping a in the logarithm), which I undertook before[1,2] is the only strict procedure. My followers have mostly written a fictitious

$$r = \sqrt{(x - \xi)^2 + a^2},$$

the distance between a point on the surface and a point on the axis, instead of $|x - \xi|$, and even used a square root expression in the first term of (5a). In doing so they have not at all increased the accuracy and in fact have made (5a) unsolvable, in a strict mathematical sense, because the derivative of the left side then will remain continuous at $x = 0$, but not of the right side. Gans, in his above mentioned papers, criticizes this inconsistency, which he quite erroneously attributes to me. The remedy which he recommends, the changing of the right side of (5a) as well, does not seem adequate. It is better to return to the original for this integral equation method.[2] (When there is no potential jump, passive antennas, and the antenna is solid, the distance $r = \sqrt{(x - \xi)^2 + a^2}$ may be used, because one can use the vanishing electric field along the antenna axis to get an integral equation, but then there are always end surfaces.)

From the exact symmetrical expression (9) for the outgoing traveling current wave it is very easy to get the potentials and field strengths at any point outside the antenna and even inside the antenna tube. We only have to apply the formula

$$\frac{1}{2\pi}\int_0^{2\pi} d\varphi \int_{-\infty}^{\infty} \frac{e^{-i\beta r + ia\xi}}{r}\, d\xi$$

$$= \begin{cases} 2e^{ia z}I_o(a\sqrt{\alpha^2 - \beta^2})K_o(\rho\sqrt{\alpha^2 - \beta^2}), & \rho > a \\ 2e^{ia z}I_o(\rho\sqrt{\alpha^2 - \beta^2})K_o(a\sqrt{\alpha^2 - \beta^2}), & \rho < a \end{cases}, \quad (14)$$

where $r = \sqrt{(x - \xi)^2 + \rho^2 + a^2 - 2a\rho\cos\varphi}$ is the distance between two points, whose cylinder co-ordinates are ρ, 0, x respective a, φ, ξ. By putting (9) in the integral expression for the vector potential (similar to the left side of (3)) we get with the help of (14) the vector potential at any point:

$$A_x = -\frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\alpha^2 - \beta^2}\cdot\frac{K_o(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho > a \tag{15a}$$

$$A_x = -\frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\alpha^2 - \beta^2}\cdot\frac{I_o(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho < a \tag{15b}$$

$$A_x = -\frac{V_o}{c}e^{-i\beta|x|}, \qquad \rho = a. \tag{15c}$$

As the scalar potential $V = jc/\beta\ (\partial A_x/\partial x)$ we get from (15):

$$V = V_o\frac{j}{\pi}\int_\Gamma \frac{\alpha e^{ia z}}{\alpha^2 - \beta^2}\cdot\frac{K_o(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho > a \tag{16a}$$

$$V = V_o\frac{j}{\pi}\int_\Gamma \frac{\alpha e^{ia z}}{\alpha^2 - \beta^2}\cdot\frac{I_o(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho < a \tag{16b}$$

$$V = -V_o\frac{x}{|x|}e^{-i\beta|x|}, \quad \rho = a. \tag{16c}$$

As the magnetic field is $B_\varphi = -\partial A_x/\partial\rho$ we also get from (15):

$$B_\varphi = -\frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\sqrt{\alpha^2 - \beta^2}}\cdot\frac{K_1(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \rho > a \tag{17a}$$

$$B_\varphi = \frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\sqrt{\alpha^2 - \beta^2}}\cdot\frac{I_1(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho < a \tag{17b}$$

$$B_\varphi = -\frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\sqrt{\alpha^2 - \beta^2}}\cdot\frac{K_1(a\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha,$$
$$\rho = a + 0 \tag{17c}$$

$$B_\varphi = \frac{V_o}{c}\cdot\frac{j\beta}{\pi}\int_\Gamma \frac{e^{ia z}}{\sqrt{\alpha^2 - \beta^2}}\cdot\frac{I_1(a\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha,$$
$$\rho = a - 0. \tag{17d}$$

The axial electric field component is

$$E_z = -\frac{jc}{\beta}\left(\frac{\partial^2 A_x}{\partial x^2} + \beta^2 A_x\right):$$

$$E_z = \frac{V_o}{\pi}\int_\Gamma e^{ia z}\frac{K_o(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho > a \tag{18a}$$

$$E_z = \frac{V_o}{\pi}\int_\Gamma e^{ia z}\frac{I_o(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho < a \tag{18b}$$

$$E_z = 0, \qquad \rho = a. \tag{18c}$$

The radial electric field component is $E_\rho = -\dfrac{\partial V}{\partial\rho}$:

$$E_\rho = V_o\frac{j}{\pi}\int_\Gamma \frac{\alpha e^{ia z}}{\sqrt{\alpha^2 - \beta^2}}\cdot\frac{K_1(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})}\, d\alpha, \quad \rho > a \tag{19a}$$

$$E_\rho = -V_o \frac{j}{\pi} \int_\Gamma \frac{\alpha e^{i\alpha x}}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{I_1(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha, \quad \rho < a$$

(19b)

$$E_\rho = V_o \frac{j}{\pi} \int_\Gamma \frac{\alpha e^{i\alpha x}}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{K_1(a\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha,$$

$$\rho = a + 0 \quad (19c)$$

$$E_\rho = -V_o \frac{j}{\pi} \int_\Gamma \frac{\alpha e^{i\alpha x}}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{I_1(a\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha,$$

$$\rho = a - 0. \quad (19d)$$

From (17c) we get the outside current; i.e., the true antenna current $I_e = B_\varphi\, 2\pi a/\mu_o$:

$$I_e = -\frac{4\pi}{Z_o} V_o \frac{ja\beta}{2\pi} \int_\Gamma \frac{e^{i\alpha x}}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{K_1(a\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha,$$

(20)

and from (17d) the inner current (waveguide current) $I_i = -B_\varphi(2\pi a/\mu_o)$:

$$I_i = -\frac{4\pi}{Z_o} V_o \frac{ja\beta}{2\pi} \int_\Gamma \frac{e^{i\alpha x}}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{I_1(a\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha. \quad (21)$$

The expression (20) is identical with (11, 12) and the sum of (20) and (21) is identical with (9). In (17d), (19d), and (21d) the integral can easily be carried out and gives a series of wave guide waves below cutoff frequency (see reference 3, eq. 35.25). The character of the potential jump $2 V_o$ in the feeding point and the direction we have given it in this investigation is clearly seen in (16c). In the following we will also give some other forms and asymptotic values for some of the quantities (15)–(21).

INTEGRAL EQUATION FOR REFLECTED ANTENNA CURRENT WAVES

After this survey of the outgoing current wave we turn to the main object, the reflected waves. The first investigation I made on them[9] was based on the principle that a standing wave expression was first obtained and then this standing current wave was dissolved into a system of traveling waves. Thus the expression (reference 9, eq. 55) for the first, second and following traveling waves was obtained. However, there have as yet only existed approximate expressions for the standing waves, consisting of a series of a limited number of terms, obtained by iteration from the integral equation (5a). Thus the traveling waves will also be known through series with a limited number of terms. This method is laborious, as already the third term in the standing current wave expression (reference 9, eq. 30) is extremely complicated. A new direct method of finding the traveling waves has been developed later.[3] It consists of splitting up (5a) into a series of integral equations, one for each traveling wave. This can be done by studying the periodicity of both sides of (5a) with respect to x as well as with respect to l. Thus the series of integral equations (reference 3, eqs. 35.61,

35.62a, b, c) for the outgoing wave, for the first, second, etc., reflected waves, is obtained. From these equations it is very easy to find the result (55) in reference 9, and even increase the number of terms; i.e., the accuracy, in the series solutions (reference 3, eqs. 35.65, 35.66).

These new integral equations also invite us to try to solve the antenna problem exactly in the form of integrals without any abbreviated series. In fact, since Levine and Schwinger[12] treated sound waves in a limited unflanged pipe with the help of Hopf's and Wiener's solution to a certain integral equation, the integral extending from 0 to ∞, it has been evident that the problem of the cylindrical antenna (with tube-shaped ends) can also be solved mathematically exactly. In my book[3] I have also given the exact solution of the problem of the first reflected wave, as it is derived from the linearized antenna integral equation (5a), when the distance l from the end to the feeding point tends to infinity. It is also quite certain that a finite l makes no real extra difficulties and that the following waves can be exactly determined as well, although formulas become more involved in these cases. It is our object in this paper to make a corresponding investigation based on the exact integral equation (3) for a tube-shaped cylindrical antenna; i.e., to obtain the exact expression for the first reflected wave, and to compare the result with the corresponding result from the linearized equation. In doing so we will get an exact knowledge of the particular role played by the antenna ends. Of course the tube-shaped ends are only one of several possibilities, but for the role of the ends the tube end can serve as a type. As I expected—I do not answer for others—the difference between the results from the exact equation (3) and the linearized equation (5a) consists in terms which have the small factor $a\beta$. For thin antennas in fact the linearized equation is the most general, it is independent of the particular form of the ends, and what is common to all antennas with differently shaped end surfaces is just what the linearized equation gives. But it is very important to know, numerically, for a special case how far the agreement between the exact solution and the solution of the linearized equation extends.

Two features can be expected to be of special interest. When the reflected traveling current wave has traveled a considerable distance from the end it is obviously of very little concern whether we have used the exact or the linearized equation. But in the beginning and especially at the very end-point of the antenna the difference must be expected to be the greatest. Therefore we will determine the exact end admittance of the reflected wave as it is derived from (3) and compare it with the end admittance derived from (5) which has already been determined.[3] Further, the open tube-end will carry a

---

[12] H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," *Phys. Rev.*, vol. 73, pp. 383–406; 1948.

small internal current, which analog to the inner current of the outgoing wave can be expected to be aperiodical in space and only represent an end-charging current. This mainly capacitive current branches off from the reflected wave current at the end of the antenna. We will determine it numerically at the end-point. This current has no correspondence in the linearized solution, where there is, however, just as was the case with the outgoing wave, a fictitious small, slow, extra wave current without physical significance.[3] Besides this, the exact shape of the field round the antenna end will be of interest.

For the deduction from (5a) of the complete system of integral equations, one for each traveling current wave, see reference 3, pages 421–425, 433–434. Although the main feature of the reasoning is based only on the true antenna waves, it can be repeated in exactly the same way even if we start with the exact standing wave equation (3), in spite of the fact that its general solution contains even the aperiodic tube waves. In this paper we will, however, limit the investigation to the first reflected wave, and, for simplicity, assume the antenna length $2l$ to be very great. The influence of l on the field and current conditions around the end apparently is very small anyway. Then we do not need the whole system of integral equations and can obtain the needed one in a simple way.

We introduce in our equations the distance z along the antenna from the end and thus have $x = z - l$, where $z = 0$, when $x = -l$. We usually need not extend z to greater values than $z < l$ and (for this part of the antenna) have $|x| = l - z$. The incoming current wave according to (9) is in our new co-ordinate:

$$I(z) = -\frac{4\pi}{Z_o} V_o \frac{j\beta}{2\pi}$$

$$\cdot \int_\Gamma \frac{e^{j\alpha(z-l)}}{(\alpha^2 - \beta^2)I_o(a\sqrt{\alpha^2 - \beta^2})K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha. \quad (22)$$

If l increases infinitely there will be no reflected wave coming from the farther end $z = 2l$, but only, besides the incoming current $I(z)$ and corresponding potential wave $V_o e^{-i\beta|l-z|}$, one reflected current wave $i(z)$ and a corresponding potential wave $V_1 e^{-i\beta z}$. Thus we get from (3), when $l \to \infty$ the integral equation

$$\int_0^\infty [I(\zeta) + i(\zeta)]d\zeta \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi$$

$$= \frac{4\pi}{Z_o}(-V_o e^{-i\beta|l-z|} + V_1 e^{-i\beta z}), \quad (23)$$

where $\zeta = \xi + l$ and $r = \sqrt{(z - \zeta)^2 + 4a^2 \sin^2 \frac{1}{2}\varphi}$. Now, if we fictively define $I(\zeta)$ as before, even for negative ζ, we have, according to (4),

$$\int_{-\infty}^\infty I(\zeta) d\zeta \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi = -\frac{4\pi}{Z_0} V_o e^{-i\beta|l-z|}.$$

Subtracting this equation from (23) we get the *integral equation for the first reflected current wave*:[3]

$$\int_{-\infty}^\infty i(\zeta) d\zeta \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi = \begin{cases} \dfrac{4\pi}{Z_o} V_1 e^{-i\beta z}, & z > 0 \\ g(z), & z < 0, \end{cases} \quad (24)$$

where $i(z)$ is, for $z > 0$, the unknown reflected current wave, and, by arbitrary definition, for $z < 0$:

$$i(z) = -I(z). \quad (25)$$

$g(z)$ is an unknown function which we need not determine. V_1 is a potential amplitude, which has to be determined. Eq. (24) is valid whether the antenna length is infinite or finite. Now we will use the simplification which comes from making $l = \infty$. The only consequence of this is that the incoming current (22) will be simplified. In fact, the expression (22) for the incoming current wave approaches, for big l,

$$I(z) = -\frac{4\pi}{Z_o} V_o e^{i\beta(z-l)}\Omega_{l-z}^{-1}, \quad (26)$$

where

$$\Omega_{l-z} = \log \frac{2(l - z)}{a^2\beta} - \gamma - j\frac{\pi}{2} + l(2\beta(l - z)). \quad (27)$$

In (26) terms of the order of magnitude Ω_{l-z}^{-3} have been neglected. Neglecting even z beside l we get a constant amplitude of the incoming current wave by substituting Ω_l for Ω_{l-z}, where $\Omega_l = \log 2l/a^2\beta - \gamma - j\pi/2$. The corresponding asymptotic expressions for the incoming wave of the vector potential and the scalar potential as derived from (15a), (16a) are (with replaced by $z - l$):

$$A_z = -\frac{V_o}{c} e^{i\beta(z-l)} - \frac{V_o}{c} 2\log \frac{a}{\rho} e^{i\beta(z-l)}\Omega_{l-z}^{-1}, \quad \rho > a \quad (28)$$

$$V = V_o e^{i\beta(z-l)} + V_o 2\log \frac{a}{\rho} e^{i\beta(z-l)}\Omega_{l-z}^{-1}, \quad \rho > a, \quad (29)$$

where l is very big in comparison to z, ρ, a. For $\rho < a$ we simply get from (15b) and (16b) the asymptotic values:

$$A_z = -\frac{V_o}{c} e^{i\beta(z-l)} \quad (30)$$

$$V = V_o e^{i\beta(z-l)}, \quad (31)$$

and in this case the approach is closer; the neglected terms are small essentially as $e^{-l/a}$.

The first terms in (28), (29) together with (30), (31) form a potential "field" of constant amplitude and constant direction. It has no physical content because the field strengths B and E as derived from it are zero everywhere. The electromagnetic potentials of a radiation field are more purely mathematical and less physical than in quasi-stationary cases. It is therefore more adequate to use the incoming current wave amplitude as a measure of the incoming wave than the potential [which we keep only in the unimportant first terms of (28)–(31)]. We put the constant

$$I = \frac{4\pi}{Z_o} V_o e^{-i\beta l}\Omega_l^{-1} \quad (32)$$

nd have for the incoming wave the asymptotic for-
aulas

$$I(z) = -Ie^{i\beta z} \qquad (33)$$

$$_z = -\frac{V_o}{c} e^{i\beta(z-l)} - \frac{\mu_o I}{2\pi} \log \frac{a}{\rho} e^{i\beta z}, \qquad \rho > a \qquad (34a)$$

$$_z = -\frac{V_o}{c} e^{i\beta(z-l)}, \qquad \rho < a \qquad (34b)$$

$$= V_o e^{i\beta(z-l)} + \frac{Z_o I}{2\pi} \log \frac{a}{\rho} e^{i\beta z}, \qquad \rho > a \qquad (35a)$$

$$V = V_o e^{i\beta(z-l)}, \qquad \rho < a \qquad (35b)$$

$$B_\varphi = -\frac{\mu_o I}{2\pi\rho} e^{i\beta z}, \qquad \rho > a \qquad (36a)$$

$$B_\varphi = 0, \qquad \rho < a. \qquad (36b)$$

The expressions (34)–(36) have an independent physical
ense only for $z > 0$. For $z < 0$; i.e., outside the antenna
nd, one cannot reasonably make a difference between
ncoming and reflected wave (or this difference would be
rtificial). In that case (34)–(36) represent a first part
f a total wave (see further below).

SOLUTION OF THE INTEGRAL EQUATION FOR THE REFLECTED CURRENT WAVE

We now introduce (33) in (25) and (24) and turn to
he exact solution of the integral equation (24). The
rocedure for solving (24) is exactly the same as that
which I have carried out before for the corresponding
inearized equation.[3] The only difference is that we have
different kernel, which now is the exact one for a tube-
haped antenna. The mathematical method is in its
main feature that of Hopf and Wiener[13] and it was first
ased on a physical tube problem (an accoustical one)
y Levine and Schwinger.[12] Our procedure is, however,
ntirely different from that of the latter. One of the
dvantages we get is that we never need a study of the
ield to get boundary conditions, but that these are
utomatically filled and all constants known. [See the
very simple deduction of the formulas (48)–(50) below].
n fact, we do not need the formulas for the potentials
nd fields (15)–(19), (34)–(36), etc. for solving the
roblem; however, we get them, as we have seen, very
asily and they are of interest in themselves.

In order to facilitate our computation we temporarily
ssume that β has a small negative imaginary term
$\beta = \beta_1 - j\epsilon$ so that β is situated in the lower complex
emi-plane. We dissolve the kernel of the integral equa-
tion (24) into a Fourier integral (reference 3, eq. 35.22):

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-i\beta r}}{r} d\varphi$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha(z-\zeta)} 2I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2}) d\alpha.$$

Even $i(\zeta)$ is expressed as a Fourier integral

$$(\zeta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha\zeta} [F_+(\alpha) + F_-(\alpha)] d\alpha, \qquad (37)$$

where

$$F_+(\alpha) = \int_0^\infty i(u) e^{-i\alpha u} du,$$

$$F_-(\alpha) = \int_{-\infty}^0 i(u) e^{-i\alpha u} du \cdot \qquad F_+(\alpha)$$

is unknown but regular in the whole *lower* complex
α-semi-plane. When α approaches $-j\infty$ we get the
limit value of $F_+(\alpha)$:

$$F_+(\alpha) \rightarrow i(0) \int_0^\infty e^{-i\alpha u} du = \frac{i(0)}{j\alpha}. \qquad (38)$$

$F_-(\alpha)$ is known because for negative u, according to
(25), (33), we have $i(u) = Ie^{i\beta u}$, and thus

$$F_-(\alpha) = I \int_{-\infty}^0 e^{i(\beta-\alpha)u} du = \frac{I}{j(\beta-\alpha)}, \qquad (39)$$

which is a regular function of α in the whole *upper* com-
plex α-semi-plane. We finally dissolve even the right
side of (24) and find it to be

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha z} [G_+(\alpha) + G_-(\alpha)] d\alpha,$$

where $G_+(\alpha)$ is known:

$$G_+(\alpha) = \int_0^\infty \frac{4\pi}{Z_o} V_1 e^{-i\beta u} e^{-i\alpha u} du = \frac{4\pi}{Z_o} \frac{V_1}{j(\beta+\alpha)}, \qquad (40)$$

which is a regular function of α in the whole *lower*
complex α-semi-plane. $G_-(\alpha)$ is unknown:

$$G_-(\alpha) = \int_{-\infty}^0 g(u) e^{-i\alpha u} du$$

but regular in the *upper* complex α-semi-plane. It
disappears at infinity essentially as $1/-j\alpha$ in this semi-
plane.

Introducing the Fourier integrals in (24) we get

$$\int_{-\infty}^{\infty} d\zeta \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\gamma\zeta} [F_+(\gamma) + F_-(\gamma)] d\gamma$$

$$\cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha(z-\zeta)} 2I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2}) d\alpha$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha z} [G_+(\alpha) + G_-(\alpha)] d\alpha.$$

Using again Fourier's integral theorem in the op-
posite direction the integrations in ζ and in γ give us
back $[F_+(\alpha) + F_-(\alpha)]$ and our integral equation attains
the final form

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\alpha z} \{ 2I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2})$$

$$\cdot [F_+(\alpha) + F_-(\alpha)] - G_+(\alpha) - G_-(\alpha) \} d\alpha = 0. \quad (41)$$

Following the Hopf-Wiener method, we write

$$2I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2}) = \frac{\varphi_1(\alpha)}{\varphi_2(\alpha)}, \qquad (42)$$

where

$$\varphi_1(\alpha) = \exp \frac{1}{2\pi j} \int_{-\infty-j\delta}^{\infty-j\delta} \frac{d\zeta}{\zeta - \alpha}$$

$$\cdot \log [2I_o(a\sqrt{\zeta^2 - \beta^2}) K_o(a\sqrt{\zeta^2 - \beta^2})] \quad (43)$$

[13] E. C. Titchmarsh, "Fourier Integrals," Oxford University
Press, Oxford, England, pp. 1–390, especially p. 339; 1937.

$$\varphi_2(\alpha) = \exp \frac{1}{2\pi j} \int_{-\infty+j\delta}^{\infty+j\delta} \frac{d\zeta}{\zeta - \alpha}$$
$$\cdot \log \left[2I_o(a\sqrt{\zeta^2 - \beta^2})K_o(a\sqrt{\zeta^2 - \beta^2})\right]. \quad (44)$$

Here δ is a positive number so small that $\delta < \epsilon$, i.e., both the integration paths in (43), (44) go between $\beta = \beta_1 - j\epsilon$ and $-\beta = -\beta_1 + j\epsilon$. Then $\varphi_1(\alpha)$ is regular and different from zero in the whole upper complex α-semi-plane, $\varphi_2(\alpha)$ in the lower semi-plane, and both these regions can be extended somewhat to the other side of the real axis provided that the border line remains within the strip $\pm\delta$. We even find

$$\varphi_1(-\alpha)\varphi_2(\alpha) = 1. \quad (45)$$

Eq. (41) now takes the form

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\alpha z} \frac{\varphi_1(\alpha)}{\alpha^2 - \beta^2} \left\{ \frac{\alpha^2 - \beta^2}{\varphi_2(\alpha)}[F_+(\alpha) + F_-(\alpha)] \right.$$
$$\left. - \frac{\alpha^2 - \beta^2}{\varphi_1(\alpha)}[G_+(\alpha) + G_-(\alpha)] \right\} d\alpha = 0. \quad (46)$$

Here $\varphi_1(\alpha)/(\alpha^2 - \beta^2)$, $(\alpha^2 - \beta^2)/\varphi_2(\alpha)$, $(\alpha^2 - \beta^2)/\varphi_1(\alpha)$, $F_-(\alpha)$ and $G_+(\alpha)$ all remain analytical within the strip $\pm\delta$. If we anticipate, as is natural, that when β is slightly complex $(\beta = \beta_1 - j\epsilon)$, $i(z)$ decreases at least as $e^{-\epsilon z}(z > 0)$, and $g(z)$ at least as $e^{\epsilon z}(z < 0)$, then even $F_+(\alpha)$ and $G_-(\alpha)$ will remain analytic within the strip. The integration path in (46) thus can be arbitrary if it remains within the strip. The necessary condition for this is that the bracket expression vanishes and therefore with respect to (39), (40)

$$P(\alpha) = \frac{\alpha^2 - \beta^2}{\varphi_2(\alpha)} F_+(\alpha) + j \frac{\alpha + \beta}{\varphi_2(\alpha)} I$$
$$= -j \frac{\alpha - \beta}{\varphi_1(\alpha)} \frac{4\pi V_1}{Z_o} + \frac{\alpha^2 - \beta^2}{\varphi_1(\alpha)} G_-(\alpha). \quad (47)$$

The left side is analytic in the lower semi-plane, the right side in the upper semi-plane; within the strip both sides are equal. Consequently they represent the same analytic function. As F_+ and G_- are bounded, $P(\alpha)$ must be a polynomial of no higher than the second degree, and as F_+ and G_- vanish at infinity it cannot even be of more than the first degree. We may thus put

$$P(\alpha) = C_1 + C_2(\alpha - \beta),$$

where C_1 and C_2 are two constants.

Putting $\alpha = \beta$ we get, from the left side of (47),

$$C_1 = j\frac{2\beta}{\varphi_2(\beta)} I.$$

If we divide (47) by $(\alpha - \beta)$ and let α go towards $-j\infty$, we get with respect to (38),

$$C_2 = \frac{\alpha + \beta}{j\alpha\varphi_2(\alpha)} i(0) + j\frac{\alpha + \beta}{(\alpha - \beta)\varphi_2(\alpha)} I = j\frac{I - i(0)}{\varphi_2(\alpha)},$$

and thus according to (25), (33):

$$C_2 = 0.$$

Hence P is a constant:

$$P = j\frac{2\beta}{\varphi_2(\beta)} I. \quad (48)$$

We determine even V_1 by using the right side in (47) where we put $\alpha = -\beta$, which gives

$$P = j\frac{2\beta}{\varphi_1(-\beta)} \frac{4\pi V_1}{Z_o},$$

or with the help of (45) and (48),

$$V_1 = \frac{Z_o}{4\pi\varphi_2^2(\beta)} I. \quad (49)$$

We can even introduce the *end admittance of the reflected wave* (when the incoming wave comes from infinity):

$$Y_\infty = \frac{I}{V_1} = \frac{4\pi}{Z_o} \varphi_2^2(\beta). \quad (50)$$

(See Fig. 1 for numerical values.)

The Fourier transform of the unknown function $i(z)$ is now known, for from (47), (48) we get

$$F_+(\alpha) + F_-(\alpha) = j\frac{2\beta}{\varphi_2(\beta)} I \frac{\varphi_2(\alpha)}{\alpha^2 - \beta^2}. \quad (51)$$

From (37) we now find the solution of the problem. The reflected current wave is

$$i(z) = I\frac{j\beta}{\pi\varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{j\alpha z}\varphi_2(\alpha)}{\alpha^2 - \beta^2} d\alpha, \quad (52a)$$

or, with respect to (42),

$$i(z) = I\frac{j\beta}{2\pi\varphi_2(\beta)}$$
$$\cdot \int_{-\infty}^{\infty} \frac{e^{j\alpha z}\varphi_1(\alpha)}{(\alpha^2 - \beta^2)I_o(a\sqrt{\alpha^2 - \beta^2})K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha. \quad (52b)$$

The factor before the integral in (52a) can also be written $V_1 j4\beta\varphi_2(\beta)/Z_o$. The properties of $\varphi_2(\alpha)$, $i(z)$ and $g(z)$ will be studied in the following.

We state that all the equations (37)–(40), (45)–(52a) are exactly the same as when the equation was linearized in Reference 3, eqs. (35.52)–(35.57). Thus (52a) and (52b) even represent the reflected current wave, as derived from the linearized equation, but with φ_1 and φ_2 defined as in (43), (44) with the difference that the bracket $[2I_oK_o]$ is replaced by a logarithmic expression, which is the limit value of the Bessel function expression if powers of $a\beta$ are suppressed. Thus there is exactly the same relation between our solution to the reflected wave integral equation (24) and the corresponding one of the linearized equation as there is between the two exact solutions (9) and (10) of the exact integral equation (4) and the linearized one (6) for the outgoing wave. It is however our task to carry out the comparison completely numerically and find out how thick the antenna may be if the linearization shall still be permitted.

It is easy to check the result (52a). For $z < 0$ we can add an infinitely big circle in the lower complex α-semi-plane to the integration path. As $\varphi_2(\alpha)$ is regular in the lower half-plane, the integrand has no other singularity within the contour than the pole $\alpha = \beta = \beta_1 - j\epsilon$ and we get for negative z the value

$$i(z) = I\frac{j\beta}{\pi\varphi_2(\beta)} (-2\pi j) \frac{\varphi_2(\beta)}{2\beta} e^{j\beta z} = Ie^{j\beta z}, \quad z < 0,$$
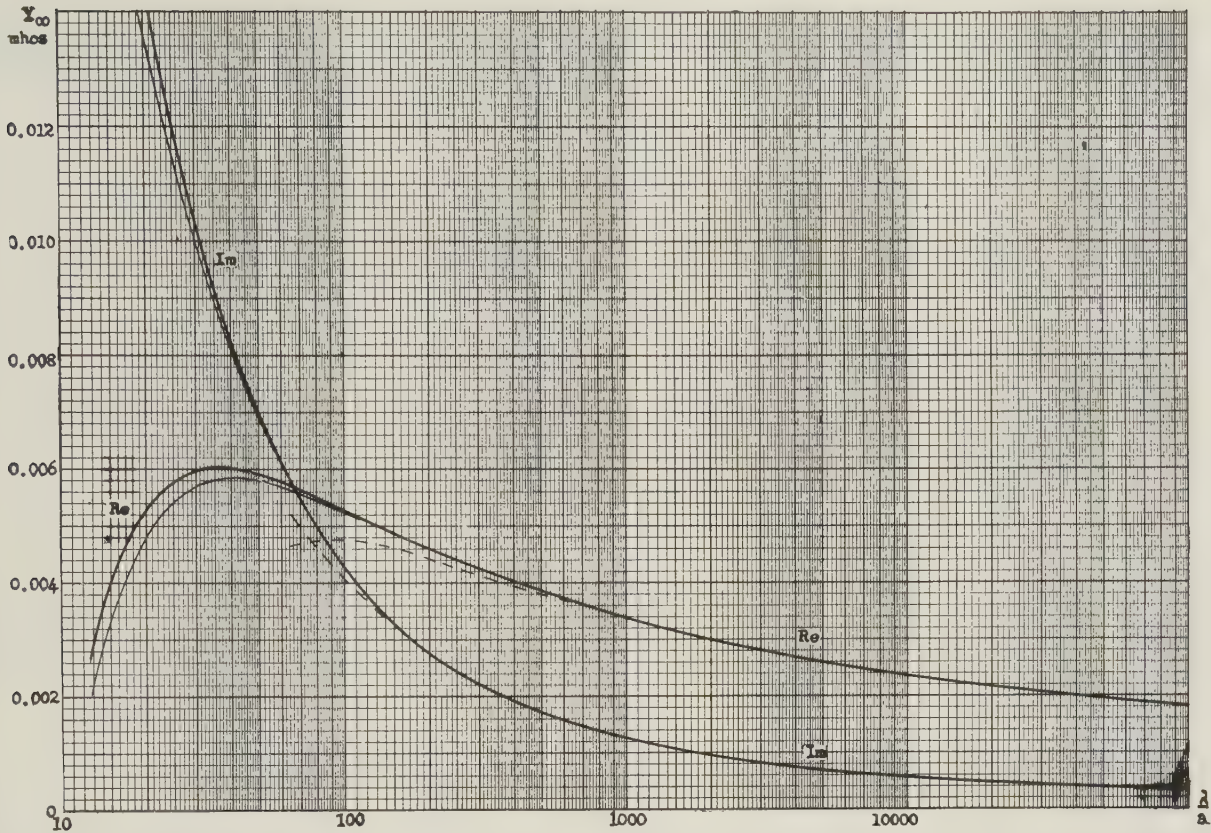
Fig. 1—Real and imaginary part of end admittance of tube-haped antenna. Solid black line indicates exact value: (67) and (72). Lighter line indicates linearized value: (68) and (69). Broken line indicates approximate series (71).

in agreement with (25), (33). This is the fictive current which we have introduced to compensate the fictive continuation of the incoming current wave:

$$i(z) + I(z) = 0, \quad z < 0.$$

For $z > 0$ we introduce (52b) in the left side of (24) and find this side of (24) to be, with respect to (14),

$$I \frac{j\beta}{\pi \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{j\alpha z}\varphi_1(\alpha)}{\alpha^2 - \beta^2} \, d\alpha, \tag{53}$$

and (for positive z) we close the integration path in the upper semi-plane, where $\varphi_1(\alpha)$ is regular. The only singularity within the contour now is the pole $\alpha = -\beta = -\beta_1 + j\epsilon$ and (53), on account of (45), (49), becomes

$$I \frac{\varphi_1(-\beta)}{\varphi_2(\beta)} e^{-j\beta z} = \frac{I}{\varphi_2{}^2(\beta)} e^{-j\beta z} = \frac{4\pi}{Z_0} V_1 e^{-j\beta z}, \quad z > 0,$$

in accordance with the right side of (24), which equation thus is fulfilled.

For negative values of z the expression (53) represents the hitherto unknown function $g(z)$ of (24).

The reflected current wave (52), with $z > 0$, consists [as was the case with the outgoing current (9)] of two parts: an outer antenna wave i_e and an inner wave guide wave i_i, below cutoff frequency, which penetrates into the tube from the end but which decreases exponentially. For thick antennas the latter is not quite unessential as it represents a current which branches

off from the main antenna current wave. It is mainly a charge current, and if the antenna has, for instance, flat ends instead of tube-shaped ones, it corresponds to the small current which charges and discharges the end surface. We can easily split (52b) into these two currents. We observe that

$$\frac{1}{I_o(a\sqrt{\alpha^2 - \beta^2}) K_o(a\sqrt{\alpha^2 - \beta^2})}$$
$$= a\sqrt{\alpha^2 - \beta^2}\left[\frac{I_1}{I_o} + \frac{K_1}{K_o}\right]_{a\sqrt{\alpha^2 - \beta^2}} \tag{54}$$

and get

$$i(z) = i_e(z) + i_i(z),$$

where

$$i_e(z) = I \frac{ja\beta}{2\pi\varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{j\alpha z}\varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{K_1(a\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha,$$
$$z > 0 \tag{55}$$

$$i_i(z) = I \frac{ja\beta}{2\pi\varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{j\alpha z}\varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{I_1(a\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} d\alpha,$$
$$z > 0. \tag{56a}$$

With residue calculus we find the latter expression to be the following series:

$$i_i(z) = I \frac{ja\beta}{\varphi_2(\beta)} \sum_{1}^{\infty} \frac{e^{-\sqrt{\xi_m{}^2 - a^2\beta^2}\, z/a}\varphi_1(j/a \sqrt{\xi_m{}^2 - a^2\beta^2})}{\sqrt{\xi_m{}^2 - a^2\beta^2}},$$

$$\tag{56b}$$

where the ξ_m are the zeros of the Bessel function $J_o(\xi)$. (56b) is a series of waveguide waves (below cutoff frequency) to which we will return later. For a corresponding series expression for the wave guide waves of the outgoing antenna see (reference 3, eq. 35.25). The linearized equation does not give these waves (56b).

THE ELECTROMAGNETIC FIELD AROUND THE ANTENNA END

We are now in the position that we can give an exact description in every detail of the field around the antenna end (if this is tube-shaped). The way which is physically most straight forward would be to use the true currents, the incoming current (33) and the reflected current (52), for $z > 0$ and express the vector potential from them in the form of an integral extending from $z = 0$ to $z = \infty$ [cf. (23)]. But it is easier, and equally correct, to extend both currents to $z = -\infty$, with the provision (25) that $i(z) + I(z) = 0$ for $z < 0$. The field from the (extended) incoming current $I(z)$ we have already in (34a), (34b), (35a), (35b), (36a), (36b). The vector potential from the (extended) reflected current (52a) is at an arbitrary point $\rho, 0, z$:

$$A_z = \frac{\mu_o}{4\pi} \mathbf{I} \frac{j\beta}{\pi \varphi_2(\beta)} \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_{-\infty}^{\infty} \frac{e^{-i\beta r}}{r} d\zeta \int_{-\infty}^{\infty} \frac{e^{i\alpha \zeta} \varphi_2(\alpha)}{\alpha^2 - \beta^2} d\alpha,$$

where $r = [(z - \zeta)^2 + \rho^2 + a^2 - 2a\rho \cos \varphi]^{1/2}$. Using (14), (42) this gives

$$A_z = \mathbf{I} \frac{\mu_o j\beta}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_1(\alpha)}{\alpha^2 - \beta^2} \cdot \frac{K_o(\rho \sqrt{\alpha^2 - \beta^2})}{K_o(a \sqrt{\alpha^2 - \beta^2})} d\alpha,$$
$$\rho > a \quad (57a)$$

$$A_z = \mathbf{I} \frac{\mu_o j\beta}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_1(\alpha)}{\alpha^2 - \beta^2} \frac{I_o(\rho \sqrt{\alpha^2 - \beta^2})}{I_o(a \sqrt{\alpha^2 - \beta^2})} d\alpha,$$
$$\rho < a. \quad (57b)$$

For $z > 0$ this represents the reflected vector potential wave, whereas the incoming vector potential wave is given by (34a), (34b). For $z < 0$, i.e., outside the antenna end, we can no longer distinguish between incoming and reflected waves but we get the *total* vector potential wave by adding (34a), (34b). In this region ($z < 0$) it is better to go back to the function φ_2 instead of φ_1. We thus get the *total* vector potential outside the antenna end:

$$A_z = -\frac{V_o}{c} e^{i\beta(z-l)} - \mathbf{I} \frac{\mu_o}{2\pi} \log \frac{a}{\rho} e^{i\beta z}$$
$$+ \mathbf{I} \frac{\mu_o j\beta}{2\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_2(\alpha)}{\alpha^2 - \beta^2}$$
$$\cdot I_o(a \sqrt{\alpha^2 - \beta^2}) K_o(\rho \sqrt{\alpha^2 - \beta^2}) d\alpha, \quad \rho > a \quad (58a)$$

$$A_z = -\frac{V_o}{c} e^{i\beta(z-l)} + \mathbf{I} \frac{\mu_o j\beta}{2\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_2(\alpha)}{\alpha^2 - \beta^2}$$
$$\cdot K_o(a \sqrt{\alpha^2 - \beta^2}) I_o(\rho \sqrt{\alpha^2 - \beta^2}) d\alpha, \quad \rho < a. \quad (58b)$$

The expressions (58a) and (58b) are in themselves valid even for $z > 0$, although we prefer to use φ_1 instead of φ_2 when $z > 0$. The reason why we always make this distinction, using φ_1 when z is positive but φ_2 when it

is negative, is that we can deform the path of integration in the upper half-plane in the first case, and in the lower one in the second. It can be shown [cf. (64), (65a)] from the integrals in (55) and (57a) that they represent, for $z > 0$, waves traveling in the direction of positive z (i.e., have a periodic factor $e^{-i\beta z}$), whereas (52a) for $z < 0$ represents waves traveling in the negative direction (periodic factor $e^{i\beta z}$).

From the vector potential we get the magnetic field $B_\varphi = -\partial A_z / \partial \rho$ and thus, using (58a,) (58b) the *total* magnetic field outside the antenna end ($z < 0$) becomes

$$B_\varphi = -\mathbf{I} \frac{\mu_o}{2\pi\rho} e^{i\beta z} + \mathbf{I} \frac{\mu_o j\beta}{2\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_2(\alpha)}{\sqrt{\alpha^2 - \beta^2}}$$
$$\cdot I_o(a \sqrt{\alpha^2 - \beta^2}) K_1(\rho \sqrt{\alpha^2 - \beta^2}) d\alpha, \quad \rho > a \quad (59a)$$

$$B_\varphi = -\mathbf{I} \frac{\mu_o j\beta}{2\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_2(\alpha)}{\sqrt{\alpha^2 - \beta^2}}$$
$$\cdot K_o(a \sqrt{\alpha^2 - \beta^2}) I_1(\rho \sqrt{\alpha^2 - \beta^2}) d\alpha, \quad \rho < a. \quad (59b)$$

These expressions for the total magnetic field are still valid even for $z > 0$, but in this case it is better to write them [using (42)] as:

$$B_\varphi = -\mathbf{I} \frac{\mu_o}{2\pi\rho} e^{i\beta z} + \mathbf{I} \frac{\mu_o j\beta}{4\pi^2 \varphi_2(\beta)}$$
$$\cdot \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{K_1(\rho \sqrt{\alpha^2 - \beta^2})}{K_o(a \sqrt{\alpha^2 - \beta^2})} d\alpha, \quad \rho > a \quad (59c)$$

$$B_\varphi = -\mathbf{I} \frac{\mu_o j\beta}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{e^{i\alpha z} \varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}} \cdot \frac{I_1(\rho \sqrt{\alpha^2 - \beta^2})}{I_o(a \sqrt{\alpha^2 - \beta^2})} d\alpha,$$
$$\rho < a. \quad (59d)$$

In the term outside the integral we recognize the field of the incoming wave (36). If we put $\rho = a$, we find from (59d) the limit value of $-2\pi a / \mu_o B_\varphi(a)$ to be equal to the inner current (56a), and from (59c) the outer limit value of $+ 2\pi a / \mu_o B_\varphi(a)$ to be equal to the sum of the incoming current (33) and the reflected outer current (55). Our interpretation of the physical significance of the two terms (55) and (56a) of the reflected current is thus justified.

It is easy to verify that outside the antenna end the magnetic field is continuous at $\rho = a$. If we subtract (59b) from (59a) and put $\rho = a$ we find from (reference 11, p. 80, eq. 20) the result zero. However, it is a lack of elegance not to have a common formula for $\rho < a$ as well as for $\rho > a$ outside the antenna end, because the limit $\rho = a$ between the two regions has no physical significance. We find a common formula by deforming the integration path in (59a) and (59b) into a double line in the lower complex α-semi-plane from $\alpha = \beta$ to $\alpha = \beta - j\infty$. If we further substitute $\sqrt{\alpha^2 - \beta^2} = -j\beta u$ we get, for $z < 0$, the common formula:

$$B_\varphi = -\mathbf{I} \frac{\mu_o j\beta}{2\pi \varphi_2(\beta)} \int_0^{\infty} \frac{e^{\beta z \sqrt{u^2 - 1}} \varphi_2(-j\beta \sqrt{u^2 - 1})}{\sqrt{u^2 - 1}}$$
$$\cdot J_o(a\beta u) J_1(\rho\beta u) du, \quad (59e)$$

valid for all ρ but only for negative z. The integration

path should avoid the singular point $u = 1$ in the upper complex semi-plane.

We finally give even the components of the electric field which are equally easy to deduce. The *total* electric field is

$$E_z = - I \frac{Z_o}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} e^{i\alpha z} \varphi_1(\alpha) \frac{K_o(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})} \, d\alpha,$$

$$\rho > a \quad (60a)$$

$$E_z = - I \frac{Z_o}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} e^{i\alpha z} \varphi_1(\alpha) \frac{I_o(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} \, d\alpha,$$

$$\rho < a \quad (60b)$$

$$E_\rho = I \frac{Z_o}{2\pi\rho} e^{i\beta z} - I \frac{Z_o j}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{\alpha e^{i\alpha z} \varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}}$$

$$\frac{K_1(\rho\sqrt{\alpha^2 - \beta^2})}{K_o(a\sqrt{\alpha^2 - \beta^2})} \, d\alpha, \qquad \rho > a. \quad (61a)$$

$$E_\rho = I \frac{Z_o j}{4\pi^2 \varphi_2(\beta)} \int_{-\infty}^{\infty} \frac{\alpha e^{i\alpha z} \varphi_1(\alpha)}{\sqrt{\alpha^2 - \beta^2}} \frac{I_1(\rho\sqrt{\alpha^2 - \beta^2})}{I_o(a\sqrt{\alpha^2 - \beta^2})} \, d\alpha,$$

$$\rho < a. \quad (61b)$$

The first term of (61a) is the incoming wave (which has no z-component). The equations (60)–(61) are valid both for positive and negative values of z, but for negative z it is better to write the common formulas for the total field outside the antenna end:

$$E_z = - I \frac{Z_o \beta}{2\pi \varphi_2(\beta)} \int_0^{\infty} \frac{u e^{\beta z \sqrt{u^2 - 1}}}{\sqrt{u^2 - 1}}$$

$$\cdot \varphi_2(-j\beta\sqrt{u^2 - 1}) J_o(a\beta u) J_o(\rho\beta u) \, du \quad (60c)$$

$$E_\rho = I \frac{Z_o \beta}{2\pi \varphi_2(\beta)} \int_0^{\infty} e^{\beta z \sqrt{u^2 - 1}}$$

$$\cdot \varphi_2(-j\beta\sqrt{u^2 - 1}) J_o(a\beta u) J_1(\rho\beta u) \, du, \quad (61c)$$

where the integration path avoids the point $u = 1$ in the upper complex u-semi-plane. The formulas (60c), (61c) are valid for all ρ but only for negative z.

Eqs. (52)–(61) give a complete mathematical solution to the problem under investigation.

INVESTIGATION OF THE ANTENNA CURRENT

The inner current (56b) is aperiodic and decreases exponentially. Its initial value at the antenna end has most interest

$$i_i(0) = I \frac{ja\beta}{\varphi_2(\beta)} \sum_1^{\infty} \frac{\varphi_1\left(\frac{j}{a}\sqrt{\xi_m^2 - a^2\beta^2}\right)}{\sqrt{\xi_m^2 - a^2\beta^2}}, \quad (62)$$

to which we will return later. It constitutes a branch-off mainly capacitive current at the end. With the help of (49) we may define an inner end admittance

$$Y_i = \frac{i_i(0)}{V_1} = \frac{4\pi}{Z_o} ja\beta\varphi_2(\beta) \sum_1^{\infty} \frac{\varphi_1\left(\frac{j}{a}\sqrt{\xi_m^2 - a^2\beta^2}\right)}{\sqrt{\xi_m^2 - a^2\beta^2}}. \quad (63)$$

The outer reflected current (55) is periodic and represents a wave: By changing the integration path through the upper complex α-semi-plane we can write it in the form

$$i_e(z) = \frac{I}{\varphi_2^2(\beta)} \psi_\infty(\beta z) e^{-i\beta z} = \frac{4\pi}{Z_o} V_1 \psi_\infty(\beta z) e^{-i\beta z}, \quad (64)$$

where

$$\psi_\infty(\beta z) = -\tfrac{1}{2}\beta\varphi_2(\beta)$$

$$\int_{-\beta}^{-\infty} \frac{e^{i(\alpha+\beta)z}\varphi_1(\alpha) \, d\alpha}{(\alpha^2 - \beta^2)K_o(a\sqrt{\alpha^2 - \beta^2})K_o(a\sqrt{\alpha^2 - \beta^2} \, l^{i\pi})} \quad (65a)$$

(here $\sqrt{\alpha^2 - \beta^2}$ is positive along the integration path). By a suitable change of variable this also becomes

$$\psi_\infty(\beta z) = \varphi_2(\beta) \frac{2j}{\pi^2}$$

$$\int_0^{\infty} \frac{e^{\beta z(i - \sqrt{u^2-1})}\varphi_1(j\beta\sqrt{u^2 - 1})}{u\sqrt{u^2 - 1} H_o^{(1)}(a\beta u) H_o^{(2)}(a\beta u)} \, du, \quad (65b)$$

where u shall avoid the point $u = 1$ in the upper complex u-semi-plane. According to (65a) the amplitude function $\psi_\infty(\beta z)$ is an aperiodic function of βz. The index ∞ is used in order to remind us of the fact that our reflected wave arises from an incoming current wave which has come from infinity. If it comes from a finite distance l the problem can also be solved but gives a more complicated formula. A finite l, however, influences the reflected wave rather slightly. It should be noted that (65a), (65b) differ only by the two factors $\varphi_2(\beta)$ and φ_1 from the corresponding formulas for the outgoing wave (reference 3, pages 411–412).

The exact numerical evaluation of $\psi_\infty(\beta z)$ from (65a) or (65b) for all values of z is unhappily very complicated. One can fall back on series expansions (see further below) which are especially good when z is not too close to 0. For $z = 0$ we have however the value of the reflected current according to (49):

$$i(0) = i_e(0) + i_i(0) = I = \frac{4\pi}{Z_o} \varphi_2^2(\beta) V_1, \quad (66)$$

and according to (50) the total end admittance for the reflected wave

$$Y_\infty = Y_e + Y_i = \frac{i(0)}{V_1} = \frac{4\pi}{Z_o} \varphi_2^2(\beta), \quad (67)$$

where $\varphi_2(\beta)$ is simpler to determine numerically. That has been done in this investigation for a series of values of $a\beta$, which give the *exact values of the end admittance of the straight cylindrical tube-shaped antenna*, shown as a diagram in Fig. 1 and Fig. 2. Instead of $a\beta$ we have plotted the ratio λ/a between wavelength in free space and radius of cross section. We have the simple relation $\lambda/a = 2\pi/a\beta$. Even Y_i has been numerically computed for a few values and as soon as sufficient computing aid is available to me I will publish a complete diagram. Thus, Y_∞, Y_e and Y_i are all known. For details of the numerical determination of φ_2 see below.

Before we proceed further we will even mention the corresponding result as derived from the linearized integral equation. In that case there is no inner current i_i. In fact, all the equations (56b), (62), (63) above
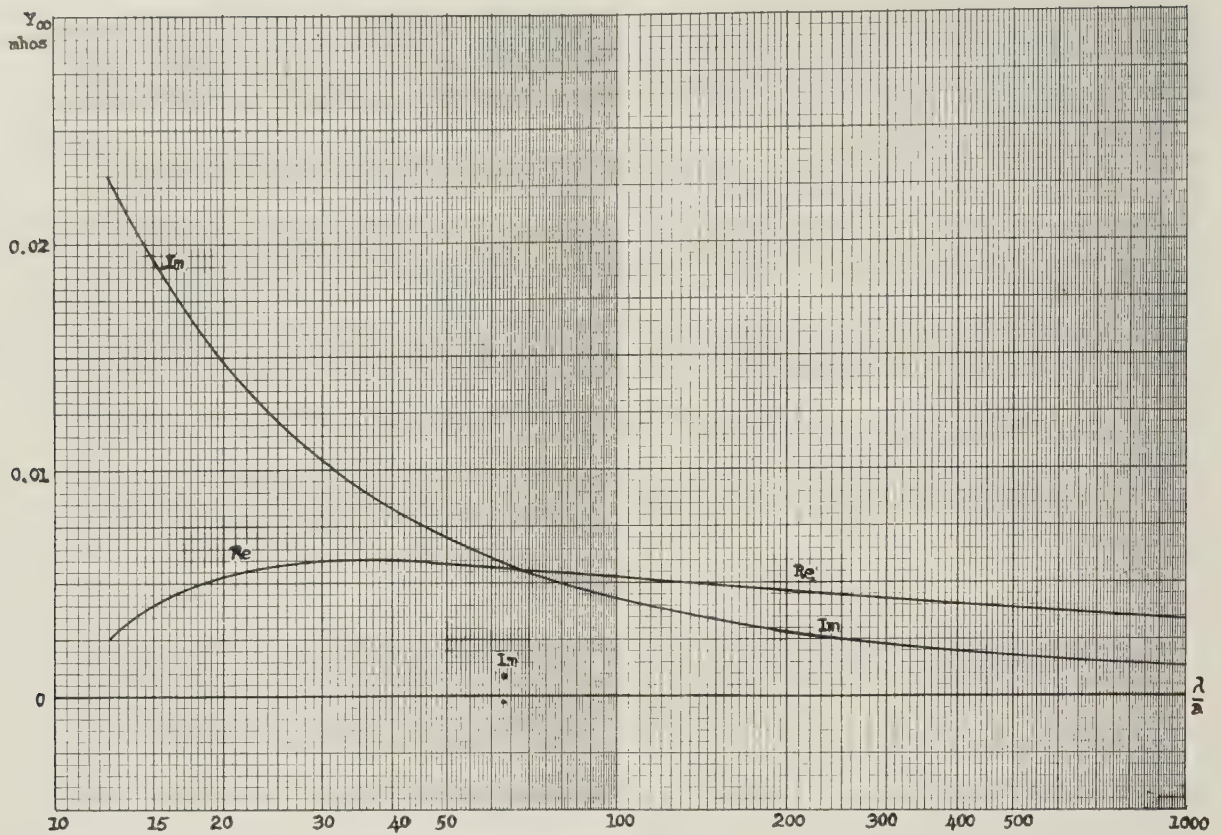
Fig. 2—Real and imaginary part of end admittance of tube-shaped antenna, exact value. Dots indicate inner end admittance (63).

contain the small factor $a\beta$, which is supposedly negligible at the linearization. Thus with linearization we have (reference 3, eq. 35.57):

$$\psi_\infty(0) = \varphi_2{}^2(\beta)$$

and

$$Y_\infty = Y_e = \frac{4\pi}{Z_0} \varphi_2{}^2(\beta) \quad \text{(linearized equation),} \quad (68)$$

where in this case

$$\varphi_2(\alpha) = \exp \frac{1}{2\pi j} \int_{-\infty+j\delta}^{\infty+j\delta} \frac{d\zeta}{\zeta - \alpha} \log \log \frac{4}{a^2(\zeta^2 - \beta^2)\gamma_1{}^2}$$

$$\text{(linearized equation).} \quad (69)$$

The function (68) has formerly been computed (reference 3, page 432, Fig. 412) and it is introduced in our Fig. 1 here also for comparison with the exact value for the tube-shaped antenna.

For arbitrary values of z the amplitude function $\psi_\infty(\beta z)$, as it follows from the linearized formula corresponding to (65b) or (65a) [or (64) and (55)], can be expanded into a rapidly decreasing series:

$$\psi_\infty(\beta z) = \Omega_z{}^{-1} + \Omega_z{}^{-3}\left[l^{11} + l^{01} - l^2 - \frac{\pi^2}{6} \right]_{2\beta z}$$

$$+ \Omega_z{}^{-4}[-3l^{111} - 2l^{101} - l^{011} + l(3l^{11} + 3l^{01} - 2l^2)$$

$$+ l_0(l^{01} - l^{11}) - 2S_3]_{2\beta z} + \cdots \text{(linearized equation),}$$

$$(70)$$

where $\Omega_z = -2\log a\beta - 2\gamma - j\pi + l_0(2\beta z) + l(2\beta z)$ and the different functions l are the amplitude functions of the iterated sine and cosine integrals of different orders as defined in (Hallen,[14] page 4; reference 9, pages 4–5; reference 3, page 435; reference 10, page 4–6). They are tabulated in (reference 10, Tables IV–VI). The parameter Ω_z increases with increasing z and the expressions within the brackets in (70) decrease, with the exception of the constant, essentially as $1/\beta z$. Thus the accuracy of (70) increases with z, and it is a very good expression. An exception is the very neighborhood of the starting point (antenna end) $z = 0$, where the series gives

$$\psi_\infty(0) = \Omega_0{}^{-1} + \frac{\pi^2}{6}\Omega_0{}^{-3} + 4S_3\Omega_0{}^{-4} + \cdots \quad (71)$$

where $S_3 = \sum_1^\infty 1/n^3 = 1.2020567$ and $\Omega_0 = -2\log a\beta - 2\gamma - j\pi$. Even this approximate value (the end admittance derived from it) is entered in Fig. 1 for comparison. I point out that generally (70) is much more exact than (71). The antenna current amplitude changes more rapidly in the beginning and there the series is least favorable. Observe also that in (70) and (71) the second negative power of Ω_z, resp Ω_0 is lacking so that

[14] E. Hallén, "Iterated sine and cosine integrals," *Trans. Roy. Inst. Tech.*, Stockholm, no. 12, pp. 1–6; 1947.

ne following terms are very small compared to the first
ne. Series corresponding to (70) for the outgoing wave
nd for reflected waves when the finite length of the
ntenna is taken into consideration have also earlier
een derived (reference 8, page 1141 and page 1145;
eference 9, eq. 55; reference 3, page 436).

The series (70) is much more easily derived directly
rom the linearized integral equation than by series
xpansion of an exact solution (reference 3, pages 433–
36). The only attempt so far to evaluate for all z-values
he exact expression (65b) for the wave amplitude of the
eflected wave on a tube-shaped antenna has been series
xpansion, but as it involves the neglect of powers
f $a\beta$, this only leads to the already known approximate
xpression (70).

It is, however, fortunate that at least the end value
or $z = 0$ is exactly numerically known (the end ad-
hittance) because when summing up the series of
raveling waves on a finite antenna, the end admittances
lay an essential role and the total current is numerically
ensitive just to this end value. So, even if at the moment
he numerical evaluation of the exact reflected current
wave along the whole antenna (from 65b) yet remains,
he fact that we know $\psi_\infty(0)$ exactly is an essential result.
Combined with (70) for those parts of the antenna which
re not quite close to the end it gives a very exact
nowledge of the antenna current everywhere.

For the numerical evaluation of the exact expression
or the antenna end admittance (50), (44) one has to
ransform the integral (44). This can be done in several
ways. If we follow the same method as in (reference 3,
age 431) for the linearized expression, we meet only two
ifferences. They are 1) the poles of the transformed
ntegral now form an infinite divergent series, and 2),
he infinite integral, before compensation, diverges.
Both difficulties can easily be overcome, and as this way
f proceeding leads mainly to Bessel functions J_o,
Y_o, of which very dense tables are available, we use this
method. Otherwise, it is easy enough to avoid the series,
but then the expression will retain functions I_o and K_o
which are not so densely tabulated.

I find (cf. reference 3, page 431) from (44):

$$\mathfrak{p}_2{}^2(\beta) = \frac{1}{-j\pi H_o{}^{(2)}(a\beta)}$$

$$\cdot \exp\left\{\frac{4}{\pi^2}\int_0^1 \frac{\log\dfrac{1+\sqrt{1-u^2}}{\sqrt{1-u^2}}}{uH_o{}^{(1)}(a\beta u)H_o{}^{(2)}(a\beta u)}\,du\right.$$

$$+\frac{4}{\pi^2}\int_1^\infty \frac{\log\dfrac{u}{\sqrt{u^2-1}}}{uH_o{}^{(1)}(a\beta u)H_o{}^{(2)}(a\beta u)}\,du$$

$$+\frac{4j}{\pi^2}\int_1^\infty \arcsin\frac{1}{u}\left[\frac{1}{uH_o{}^{(1)}(a\beta u)H_o{}^{(2)}(a\beta u)} - \frac{\pi a\beta}{2}\right]du$$

$$+2j\frac{a\beta}{\pi}\sum_1^\infty\left[\frac{1}{m} - \frac{\pi}{a\beta}\operatorname{arctg}\frac{a\beta}{\sqrt{\xi_m{}^2-a^2\beta^2}}\right]$$

$$+2j\frac{a\beta}{\pi}\left[1-\gamma-\frac{\pi}{2}+\log\frac{2\pi}{a\beta}\right]\right\}. \quad (72)$$

Except for the compensating last bracket term in the
last infinite integral and the last terms outside the
integrals, which all contain the factor $a\beta$, one recognizes
in this expression the corresponding linearized one
(reference 3, eq. 35.58c) if as usual the Hankel functions
are substituted by their first logarithmical part with
suppressing of powers of $a\beta$. The integrals in (72) are
easily converted into others with finite integrands and
finite limits (cf. reference 3, eq. 35.58d). The exact
value of the end admittance of the tube shaped antenna
(50) as it follows from (72) is shown in Fig. 1, together
with the corresponding linearized (68), (69) and the series
value as derived from (71). Fig. 2 gives a part of the
exact curve in another scale.

For the numerical evaluation of the end impedance
of the inner current (63) we proceed in the same way,
although the formula now differs because of the imagi-
nary value of α in φ_1 in (63). For summation of the series
we need an asymptotic formula for

$$\varphi_1\left(\frac{j}{a}\sqrt{\xi_m{}^2-a^2\beta^2}\right)$$

for higher values of m. This asymptotic value is found
to be

$$\varphi_1\left(\frac{j}{a}\sqrt{\xi_m{}^2-a^2\beta^2}\right) = \frac{1}{\sqrt{\xi_m}}\left(\cos\frac{k_{a\beta}}{\xi_m} - j\sin\frac{k_{a\beta}}{\xi_m}\right) \quad (73)$$

where

$$k_{a\beta} = \frac{a\beta}{\pi}\left\{\operatorname{arctg}\frac{J_o(a\beta)}{-Y_o(a\beta)}\right.$$

$$\left. -\frac{2}{\pi}\int_0^1 \frac{u\,du}{(1+\sqrt{1-u^2})H_o{}^{(1)}(a\beta u)H_o{}^{(2)}(a\beta u)}\right\}.$$

ξ_m is defined by $J_o(\xi_m) = 0$.

The curves, shown in Fig. 1 and Fig. 2, speak for
themselves. When the ratio of wave length to radius of
cross section is more than 80–100 there is no visible
difference whatsoever whether we use my old integral
equation for thin antennas (the linearized equation) or
the exact equation for tube-shaped antenna.

Propagation in Circular Waveguides Filled With Gyromagnetic Material

L. R. WALKER AND H. SUHL†

Summary—Using a specific form for the dependence of the permeability tensor components of a ferromagnetic medium on frequency and magnetizing field, the characteristic equation for the propagation constant in circular waveguide is written down. A method for discussing the complete mode spectrum of this equation is outlined. The general behavior of the spectrum is discussed.

SINCE THE properties of a gyromagnetic medium are themselves strongly dependent upon frequency and magnetizing field, the mode spectrum of a waveguide containing such materials is inevitably an involved function of these variables. For a circularly cylindrical waveguide, completely filled with homogeneous gyromagnetic material and axially magnetized, one can discuss the spectrum rather exhaustively if specific analytic expressions are chosen to describe the material properties.[1] In the ferromagnetic case (to which this paper will be limited) where one has a permeability tensor of the form

$$\begin{vmatrix} \mu & j\kappa & 0 \\ -j\kappa & \mu & 0 \\ 0 & 0 & \mu_o \end{vmatrix},$$

the axis of magnetization being in the z-direction, one may use the formulas first given by Polder for μ and κ. These are based upon small-signal solutions of the Landau-Lifshitz equations of motion. They are

$$\mu = \mu_o \left[1 - \frac{p\sigma}{1 - \sigma^2} \right]$$

$$\kappa = \mu_o \frac{p}{1 - \sigma^2},$$

where the dimensionless parameters σ and p are given by

$$\sigma = \frac{|\gamma| H_o}{\omega} \qquad p = \frac{|\gamma| M_o}{\mu_o \omega},$$

with

   γ = gyromagnetic ratio,
   H_o = magnetizing field,
   M_o = saturation magnetization,
   ω = applied angular frequency.

These expressions are simple and adequately representative of the qualitative behavior of actual materials. If the damped Landau-Lifshitz equation is used, σ is replaced by $\sigma + j\alpha$ sgn p, where α is a small quantity. The effects of loss on propagation may clearly be examined at the end by expanding in powers of α.

———————————
   † Bell Telephone Laboratories, Inc., Murray Hill, N. J.
   [1] H. Suhl and L. R. Walker, "Topics in guided wave propagation through gyromagnetic media," *Bell Sys. Tech. Jour.*, vol. 33, pp. 579–659; May, 1954; and *Bell Sys. Tech. Jour.*, vol. 33, pp. 1133–1194; September, 1954.

Fig. 1 shows how μ and κ vary with positive σ for various magnetizations, p. μ is an even and κ an odd function of σ. One notes that μ goes through zero at $\sigma = \pm\sigma_o = \pm \sqrt{(p^2/4) + 1} - p/2$ and that μ and have poles at $\sigma = \pm 1$. The points, $\pm\sigma_o$, ± 1 and zer divide the range of σ into six regions in each of which μ and κ maintain their sign relationships.
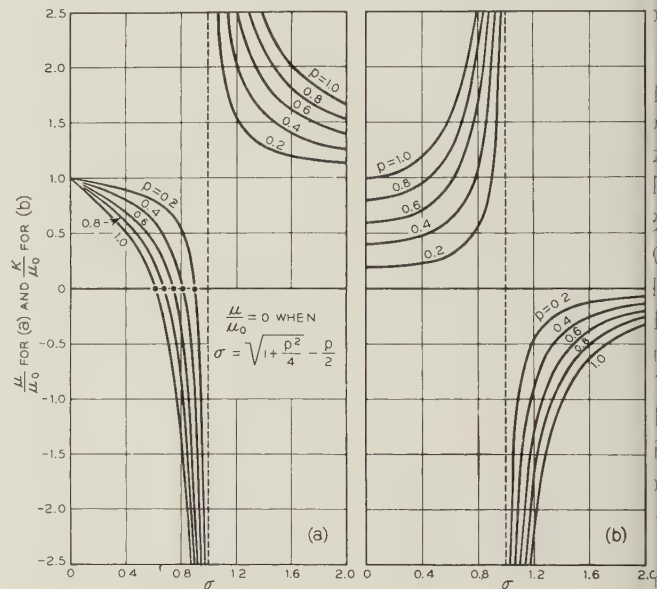
Fig. 1—The relative permeabilities μ/μ_o and k/μ_o vs σ.

One is well aware, solutions of the field equation in a gyromagnetic medium may be expressed in terms of two scalar functions, each of which satisfies a second order partial differential equation. If these two function are $f_{1,2}$, we have, if the z-dependence of all functions $e^{-i\beta z}$,

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + x_1{}^2,{}_2(\sigma, p, \beta) \right] f_{1,2} = 0.$$

At the boundary of any cylindrical wave guide we hav

$$\frac{1}{x_1{}^2} \left[\frac{j}{\lambda_1} \frac{\partial f_1}{\partial v} - \frac{\partial f_1}{\partial \sigma} \right] = \frac{1}{x_2{}^2} \left[\frac{j}{\lambda_2} \frac{\partial f_2}{\partial v} - \frac{\partial f_2}{\partial \sigma} \right]$$

and

$$f_1 = f_2$$

where $\partial/\partial v$ and $\partial/\partial \sigma$ denote differentiation normal t and along the boundary. The λ's are roots of

$$\lambda^2 - [p + \sigma(1 - \beta^2)]\lambda - \beta^2 = 0 \text{ and } x^2 = \frac{1 - \lambda^2}{1 - \sigma\lambda}.$$

For circular waveguide these boundary conditions require that the angular dependence of $f_{1,2}$ be of the form $e^{in\Phi}$, and their radial dependence is then given by th nth-order Bessel function. The boundary condition are now combined to give

$$\frac{1}{x_1^2}\left[\frac{F_n(x_1 r_0)}{\lambda_1} - n\right] = \frac{1}{x_2^2}\left[\frac{F_n(x_2 r_0)}{\lambda_2} - n\right],$$

where r_0 is the guide radius in units, $1/\omega\sqrt{\mu_0\epsilon}$, and

$$F_n(x) = x\,\frac{J_n{}'(x)}{J_n(x)}.$$

This may be rewritten as

$$G(\lambda_1,\,\sigma,\,p) = G(\lambda_2,\,\sigma,\,p) \qquad (1)$$

where, expressing x in terms of λ,

$$G(\lambda,\,\sigma,\,p) = \frac{1-\sigma\lambda}{1-\lambda^2}\left[\frac{1}{\lambda}F_n\left(r_0\sqrt{\frac{1-\lambda^2}{1-\sigma\lambda}}\right) - n\right].$$

The quadratic for λ may be decomposed into two equations

$$\lambda_1 + \lambda_2 - \sigma\lambda_1\lambda_2 = p + \sigma \qquad (2)$$

and

$$\lambda_1\lambda_2 = -\beta^2. \qquad (3)$$

Eqs. (1) and (2) may now determine for a given r_0, p and σ, a pair of values of λ, one positive (λ_1) and one negative (λ_2), which by virtue of (3) then give a positive value for β^2.

Since the transformation $\sigma \to -\sigma$, $p \to -p$, $\lambda_1 \to \lambda_2$, $-\lambda_1 \to -\lambda_2$, $n \to -n$, leaves (1)–(3) unchanged, both polarizations, $\pm n$, may be studied together by considering all σ values with σ/p positive. We shall discuss here only $n = \pm 1$.

We would like to say a little about the method of solution of these equations. Since we are concerned here primarily with the structure of the mode spectrum our problem is to keep track of the modes, to be sure that none is missed and to see how each is affected by changes in the parameters. This task is best carried out graphically: exact computation could always follow.

The function G is infinite on the lines $\lambda = 0$ and $\lambda = -1$ and on the family of curves, $1 - \lambda^2/1 - \sigma\lambda = j_n^2/r_0^2$, where the j_n's are zeros of J_n; it is zero on another infinite family which can be represented analytically. It is thus possible to dissect the $\lambda - \sigma$ plane into an infinite number of regions of positive and negative G and within these regions to draw qualitatively the constant G contours. One has also analytic information about the value of G for large λ and σ. $G(\lambda,\,\sigma)$ may then be considered to be known in the large. Further, the qualitative changes in its topography as r_0 changes can be readily followed. Since $\lambda_1 > 0$ and $\lambda_2 < 0$, (2) may be thought of as a mapping of one λ — halfplane upon the other. By considering how the infinity and zero lines in one of these half planes are thus transformed on to the other, it is possible to superimpose $G\{\lambda,\,\sigma,\,r_0\}$ and $G\{T(\lambda),\,\sigma,\,r_0\}$, where T indicates the transform of (2).

Fig. 2 shows how this works out in practice. It shows the $\lambda > 0$, $\sigma > 0$ quadrant and some lines I and O on which $G(\lambda,\,\sigma)$ is infinite or zero. The $\lambda < 0$, $\sigma > 0$ quadrant transforms into the region bounded by O_c and $(I_A')_T$ and the lines I' and O' are transforms of infinity and zero curves in the left half plane. Regions

in which $G(\lambda)$ and $G\{T(\lambda)\}$ have unlike signs clearly cannot give solutions and these are shaded. Solution curves (loci of λ with varying σ) must then lie in the unshaded regions and it is a straightforward matter to follow their course. The intersections of I-curves and of O-curves, analytic expansions near these points and the asymptotic values of the G-function all yield exact, if fragmentary, analytic information about the solution curves.
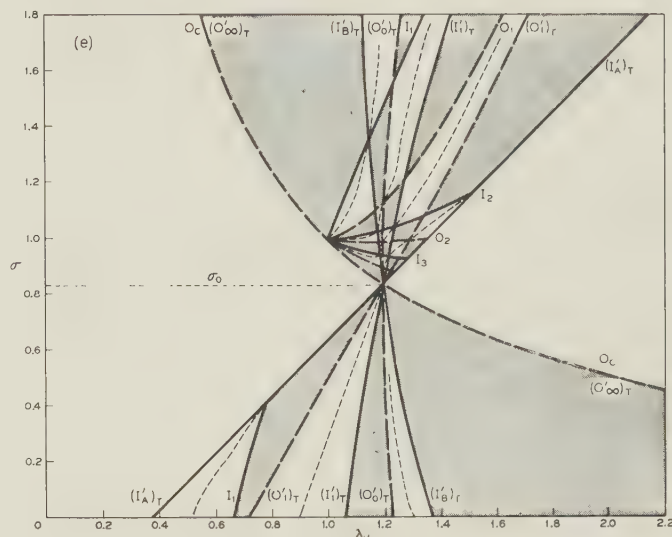
Fig. 2—The mapping of $G\{T(\lambda),\,\sigma\}$ on $G(\lambda,\,\sigma)$ for $\sigma > 0$. r_0 large enough for the guide to propagate TE_{11}-, TM_{11}-, TE_{12}-modes in the isotropic case.

In Fig. 2 the guide is sufficiently large to propagate, in the isotropic case, the TE_{11}, TM_{11} and TE_{12} modes. This is reflected in the fact that 3 solution curves are seen to persist for very large σ. The point, $\lambda = 1$, $\sigma = 1$, is the origin of an infinite number of solution curves, none passing below $\sigma = \sigma_0$ and of which all but three terminate at a finite σ. In the region $0 < \sigma < \sigma_0$ there are again three curves, one terminating at a σ-value below σ_0, two persisting right up to $\sigma = \sigma_0$.
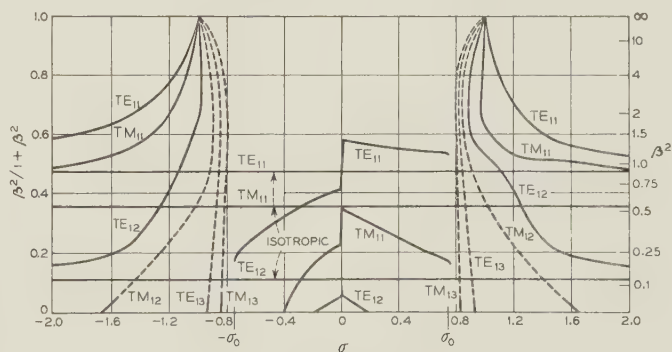
Fig. 3—The course of the fully developed modes (solid lines) and some of the incipient modes (dotted lines) as a function of σ for $|p| = 0.6$. r_0 is the same as it was in Fig. 2.

In Fig. 3, drawn for the same radius as Fig. 2, $\beta^2/1 + \beta^2$ is shown for several modes as a function of σ. There is a nonlinear β^2 scale at the right. The value of p is 0.6. Labels TE_{11}, TM_{11}, TE_{12} and so forth are

attached to these curves because it is always possible to trace a generic connection between the modes so labelled and the corresponding isotropic ones to which the name can be properly given. A number of features may be pointed out. The two polarizations are split even at $\sigma = 0$ because the medium is supposed saturated in a minute applied field. We note three types of cutoff about which more will be said later. Thus the TE_{11} is cut off at $\pm \sigma_o$ with β^2 still finite—this we call type II. TM_{11}, on the other hand is cut off at σ_o for one polarization, but cuts off at $\sigma = -1 + |p|$ and $\beta^2 = 0$ for the other polarization—this is type III. The two TE_{12} modes cut off at $\beta^2 = 0$ with a more obscure value of σ—this is type I. Above $|\sigma| = |\sigma_o|$ we see the infinite number of modes originating at $|\sigma| = 1$ with $\beta^2 = \infty$, and that three of them persist for large $|\sigma|$ with β^2 tending to the isotropic value. All the others cut off at $\beta^2 = 0$ (type I). These cutoff points move to larger σ as r_o increases and when the latter passes through a zero of J_1 or $J_1{}'$ one cutoff for each polarization goes to infinite $|\sigma|$ allowing the mode to escape. It should also be remarked that as r_o passes these critical points a new branch appears near $\sigma = 0$ and that the cutoff points of some of the modes in $|\sigma| < |\sigma_o|$ may jump discontinuously. The double-valued nature of the solutions just above $|\sigma| = |\sigma_o|$ should also be observed. One also sees that in general both polarizations exhibit the same sort of spectrum.

It is of some interest to look at the three types of cutoff point. In type I, where $\beta^2 = 0$, the value of σ is complicated but may be expressed with p and r_o in parametric form. There the E_z field goes to zero; the field is pure TM with an effective permeability, $\mu - \kappa^2/\mu$. All of the incipient modes and some below $|\sigma_o|$ are of this type. Type II occurs specifically at $|\sigma| = |\sigma_o|$ or $\mu = 0$ and here β^2 is not zero. One of the x's goes to an infinite imaginary value and the associated fields cling closely to the guide wall. With this are associated unlimited longitudinal wall currents. Finally type III with $\beta^2 = 0$ occurs only for $\mu = -\kappa$. The field is to lowest order a uniform, transverse H; H_z, E_θ, E_r are of first order in $\sigma - \sigma$ (cut off), while E_z is of the second order. It should be emphasized that any type of mode whether TE or TM in the isotropic limit, may possess any one of these types of cutoff and indeed may change from one type to another with a change in guide size.

There is a certain regularity about the mode spectrum which can be brought out in two ways. If one considers what happens to the modes in a guide of very large radius a rather curious thing appears. For $|\sigma_o| < |\sigma| < 1$, where μ is negative there are no propagating modes. For $\sigma < -1$ or $0 < \sigma < \sigma_o$ where μ and κ have the same sign, all modes tend to have the free space β for the same polarization as the circular mode considered. But, for $\sigma > 1$ or $-\sigma_o < \sigma < 0$, all modes tend to the free space β for the opposite polarization to the circular mode. In this region μ is positive, but κ negative. Now, if we calculate from the field components the amplitudes of the left and right hand polarization components of the transverse E and H fields in the circular guide a similar picture emerges, at least for the TE_{11} mode. When μ and κ have the same sign we find for either polarization a field pattern very closely related to the isotropic pattern, although we are dealing with quite different magnetic field ranges in the two cases. Similarly when μ and κ have opposite sign the field analysis gives similar patterns for each polarization, but each is quite different from the isotropic pattern.

The Low-Frequency Problem in the Design of Microwave Gyrators and Associated Elements

C. L. HOGAN†

Summary—The introduction of ferrite microwave circuit elements has allowed considerable simplification in the realization of many system functions. However, to date practical low loss ferrite devices have not been built to operate at frequencies below 3,000 mc. Many problems arise when one attempts to build devices to operate below this frequency. Some of these problems arise from the fact that mechanisms of loss occur in the ferrites at lower frequencies which are negligible at the higher microwave frequencies. In addition, at frequencies below 1,000 mc, one can seldom neglect the existence of internal anisotropy fields in the ferrite materials. The most fundamental limitation to the operation of ferrite devices at very low microwave frequencies, however, is that one is approaching the relaxation frequency for ferromagnetic resonance, and as a result the performance of all ferrite microwave devices must deteriorate at sufficiently low frequencies, regardless of whether one assumes a ferrite whose other properties are ideal. All these problems are discussed and quantitative expressions are obtained for the ultimate low-frequency limitation of ferrite isolators, circulators, and microwave gyrators.

INTRODUCTION

THE PROPERTIES of nonreciprocal microwave circuit elements are based upon the fact that the permeability which a magnetized ferrite presents to a positive circularly polarized wave is different from that which it presents to a negative circularly polarized wave.[1] As a result of this fact, the phase velocity and attenuation of a positive circularly polarized wave is different from that of a negative circularly polarized wave when propagating through a magnetized ferrite. As is well known by now, the exact solution of guided wave problems through gyromagnetic media is often extremely difficult, and one usually cannot find a general solution to most problems but is forced to find a numerical solution which is valid for a particular set of parameters. Such solutions are particularly useless to use in order to obtain information about the low frequency limit of operation of ferrite devices. However, microwave engineers are accustomed to predicting the operating characteristics of loaded waveguides from the intrinsic properties of the materials which are used to load the structure. The intrinsic properties of the materials are simply related to the behavior of infinite plane waves propagating through the material. Thus it follows that, in general, one can study the behavior of infinite plane waves propagating through magnetized ferrites, and from this analysis derive quantitative relations for the operation of guided wave devices. As will be shown later, one must apply the results of infinite plane wave analysis to waveguide devices with discretion; but as one learns to use the proper discretion according to the particular problem, the infinite plane wave analysis becomes extremely useful for predicting the ultimate behavior of most ferrite devices.[2] For this reason we shall examine in some detail in this paper the behavior of infinite plane waves propagating through magnetized ferrites, since this is a problem which can be easily solved.

Now the important property of a magnetized ferrite can be easily described in terms of the effective relative permeability which it presents to positive and negative circularly polarized magnetic fields. These permeabilities are shown in Fig. 1 and Fig. 2 for a particular set of parameters. The equations of these curves (these are relative permeabilities) are:[3]

$$\mu_+' = 1 + \frac{4\pi M_z \gamma}{\gamma H_z - \omega} \tag{1}$$

$$\mu_-' = 1 + \frac{4\pi M_z \gamma}{\gamma H_z + \omega} \tag{2}$$

$$\mu_+'' = \frac{4\pi\lambda\omega(\gamma^2 H_o^2 + \omega^2 + 2\omega\gamma H_z)}{(\gamma^2 H_o^2 - \omega^2)^2 + \frac{4\omega^2\lambda^2 H_z^2}{M_z^2}} \tag{3}$$

$$\mu_-'' = \frac{4\pi\lambda\omega(\gamma^2 H_o^2 + \omega^2 - 2\omega\gamma H_z)}{(\gamma^2 H_o^2 - \omega^2)^2 + \frac{4\omega^2\lambda^2 H_z^2}{M_z^2}}, \tag{4}$$

where

  γ = gyromagnetic ratio of electrons contributing to ferromagnetism of material (assumed to be a positive quantity in above equations),

  H_z = dc magnetic field applied to material (oersteds),

  ω = angular frequency of wave,

  M_z = Z-component of magnetization of material (gauss),

  λ = parameter which measures magnitude of damping of electrons in precessional motion (relaxation frequency in ferromagnetic resonance),

† Gordon McKay Laboratory of Applied Science, Harvard University, Cambridge, Mass.
[1] It is inconvenient to speak in terms of right- and left-handed circularly polarized waves according to the IRE standards when dealing with the subject of magnetized ferrites, since the effective permeability which the ferrite presents to a right-handed circularly polarized wave would then depend upon the direction of propagation of the wave. A positive circularly polarized wave is defined as one which is rotating in the direction of the positive current which produces the magnetic field to magnetize the ferrite. This definition is independent of the direction of propagation of the wave, and the ferrite presents a definite effective permeability to a positive circularly polarized wave which is independent of the direction of propagation.

[2] The author has previously shown, for instance, (C. L. Hogan, "The Microwave Gyrator," *Bell Sys. Tech. Jour.*, vol. 31, pp. 1–31, January, 1952.) that the rotation of the plane of polarization of an infinite plane wave predicts quantitatively the rotation in a completely filled round waveguide which is sufficiently far from cutoff for both circular polarizations.

[3] C. L. Hogan, "Ferromagnetic Faraday effect at microwave frequencies," *Rev. Mod. Phys.*, vol. 25, pp. 253–263, 1953.
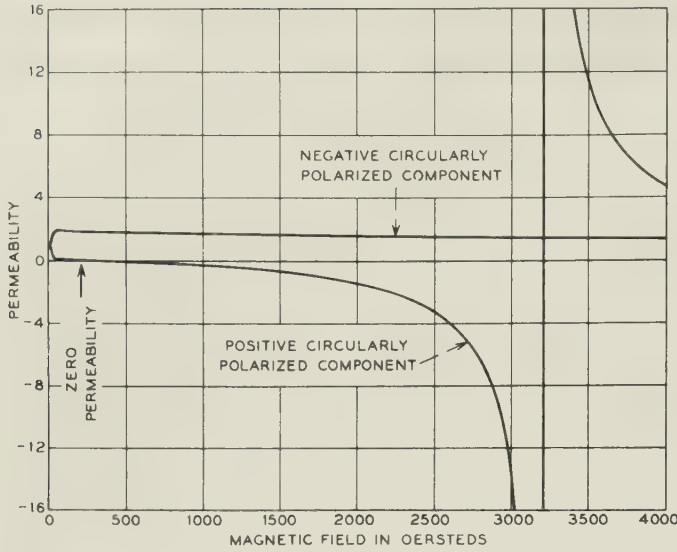
Fig. 1
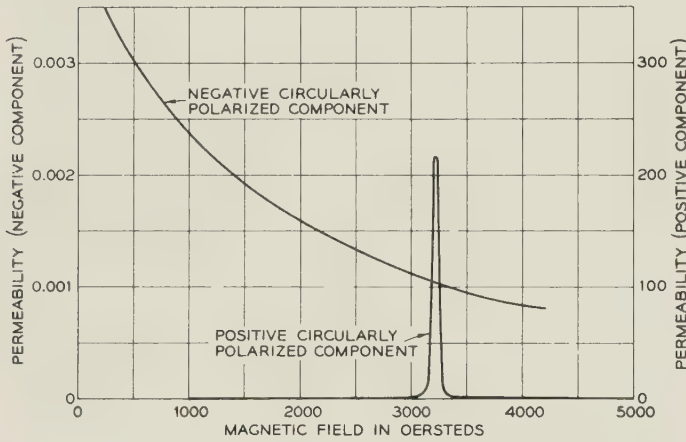
Fig. 2

$$H_o = H_z \left[1 + \frac{\lambda^2}{\gamma^2 M^2} \right]^{1/2}.$$

Now if an infinite plane wave is propagated through the above material in the Z-direction, and if it is assumed that all losses are zero except those arising from the imaginary part of the permeabilities described by (3) and (4), then the attenuation and phase constant of the waves are given by

$$\alpha_\pm = \omega \sqrt{\epsilon \mu_o \mu_r^\pm}, \tag{5}$$

$$\beta_\pm = \omega \sqrt{\epsilon \mu_o \mu_L^\pm}, \tag{6}$$

where

$$\mu_r^\pm = \frac{\sqrt{(\mu_\pm')^2 + (\mu_\pm'')^2} - \mu_\pm'}{2} \tag{7}$$

$$\mu_L^\pm = \frac{\sqrt{(\mu_\pm')^2 + (\mu_\pm'')^2} + \mu_\pm'}{2} \tag{8}$$

Provided that $| \mu_\pm' |^2 \gg | \mu_\pm'' |^2$, $\mu_r^\pm$ can be written in the particularly simple form

$$\mu_r^\pm = \frac{(\mu_\pm'')^2}{4 \mu_\pm'}. \tag{9}$$

Now the loss which is suffered by an infinite plane wave propagating through any medium is related to the attenuation constant by the following relation:

$$\frac{db \text{ loss}}{\text{meter path}} = L = 8.7\alpha; \tag{10}$$

or substituting from above:

$$L_\pm = 8.7\omega \sqrt{\epsilon \mu_o \mu_r^\pm} \tag{11}$$

and in those regions of Figs. 1 and 2 in which

$$| \mu_\pm' |^2 \gg | \mu_\pm'' |^2, \tag{12}$$

(11) can be written as

$$L_\pm = \frac{8.7\omega \sqrt{\epsilon \mu_o \mu_\pm''}}{2 \sqrt{\mu_\pm'}}. \tag{13}$$

Eq. (12) is valid everywhere for the negative circularly polarized component, provided other magnetic losses (i.e., domain wall losses or natural ferromagnetic resonance of demagnetized sample), in addition to those arising from the imaginary permeability of (4), do not occur. From Figs 1 and 2 it is evident that (12) is valid for the positive component everywhere, except in the very close vicinity of the zero permeability region of Fig. 1 and in the vicinity of the absorption line shown in Fig. 2. In these regions, where

$$| \mu_+'' |^2 \gg | \mu_+' |^2, \tag{14}$$

then (11) takes the form

$$L_+ = 8.7\omega \frac{\sqrt{\epsilon \mu_o \mu_+''}}{2}. \tag{15}$$

Low-Frequency Limitation of Ferrite Isolator

A device of primary importance to microwave techniques is a nonreciprocal attenuator or isolator which depends upon ferromagnetic resonance loss in the ferrite for loss in one direction through the device. This device can be built by propagating a plane polarized wave along the Z-axis and putting a quarter-wave plate just before the ferrite medium and just after it. Thus, while propagating through the ferrite in one direction, the wave will be a negative circularly polarized wave; and while propagating through the ferrite in the opposite direction, the wave will be a positive circularly polarized wave. The front to back ratio of this isolator in db will be the ratio of (15) and (13), if the ferrite is biased to the region of the absorption line (i.e., $\gamma H_o = \omega$). Thus the front to back ratio is

$$R = \frac{L_+}{L_-} = \frac{\sqrt{2} \sqrt{\mu_+'' \mu_-'}}{\mu_-''}, \tag{16}$$

when $\gamma H_o = \omega$,

$$\mu_+'' \simeq \frac{4\pi M_z^2 \gamma^2}{\lambda \omega} \tag{17}$$

$$\mu_-'' \simeq \frac{\pi \lambda}{\gamma H_z} \simeq \frac{\pi \lambda}{\omega} \tag{18}$$

$$\mu_-' \simeq 1 + \frac{2\pi M_z \gamma}{\omega}. \tag{19}$$

Now in general at the lower frequencies (less than 1,000 mc),

$$\frac{2\pi M_z\gamma}{\omega} > 1; \qquad (20)$$

so that one can write approximately

$$\mu_-' \simeq \frac{2\pi M_z\gamma}{\omega} \qquad (21)$$

and (16) becomes

$$R = \sqrt{\frac{16M^3\gamma^3}{\lambda^3}}. \qquad (22)$$

Now if the half-width of the resonance absorption line (shown in Fig. 2) is denoted by ΔH, then it can be easily shown that

$$\lambda = \frac{\gamma M \Delta H}{2H_r}, \qquad (23)$$

where

H_r = field required for resonance,

ΔH = difference between magnetic fields at two points where absorption line is equal to one half its maximum value.

Thus (22) becomes

$$R = \sqrt{128\left(\frac{H_r}{\Delta H}\right)^3}, \qquad (24)$$

or

$$H_r = \frac{R^{2/3}}{5.0} \Delta H.$$

If now a front to back ratio of 10 is required (i.e., 1 db forward loss to 10 db reverse loss), then

$$H_r \geq .92\Delta H, \qquad (25)$$

or the field necessary for resonance must be larger than .92 times the line width.

Little is known about the resonance line width and how it should depend upon frequency. However, line widths as narrow as 50 oersteds have been measured at all frequencies from 500 mc to 24,000 mc. Thus it seems from experimental data available today that one can obtain a line which is as narrow as 50 oersteds but not much narrower. This minimum width appears to be independent of the frequency, even though any one particular sample of ferrite has a line width which depends upon the frequency at which it is measured. Thus one can say that, theoretically, an isolator can be built which has a front to back ratio of 10, provided that the magnetic field required for resonance is greater than 45 oersteds. In other words, the isolator must operate at a frequency greater than

$$f \geq 2.8(45) = 130 \text{ mc.} \qquad (26)$$

Below this frequency, the isolator will rapidly deteriorate unless line widths less than 50 oersteds can be obtained. Since this at present seems impractical, one can say that the lower practical limit for a resonance absorption type isolator is around 100–200 mc.

To be sure, the above calculation was made for an infinite plane wave propagating through a semi-infinite medium. However, it can be easily shown that all waveguide geometries which have been tried to date lead to a low-frequency limit which is higher than this, except for the geometry of thin disks placed coaxially in a round waveguide. Most geometries lead to difficulties, because the ferrite is at resonance before it becomes completely magnetized when it is used at low frequencies (below 1,000 mc). Thus, for instance, if one were to build a device such as that described above by placing a long, slender cylinder of ferrite in a round waveguide, it would be necessary that

$$\frac{2\pi M_s\gamma}{\omega} < 1, \qquad (27)$$

in order that the ferrite could be magnetized before the region of ferromagnetic resonance is reached. It will be seen that this condition contradicts the assumption made in (20), from which (22) was derived. One must use more exact formulas for this geometry, and these formulas lead to results which are not as promising as the infinite plane wave analysis.

COMPLICATIONS AT LOW FREQUENCIES ARISING FROM USE OF NONIDEALIZED FERRITES

In order to calculate the ultimate low-frequency limit of the above device, ideal properties were assumed for the ferrite material under consideration. It is well to state clearly what assumptions are being made and whether one can actually realize them. In the first place, dielectric losses were completely neglected. As a result of work performed at the Bell Telephone Laboratories and reported in the paper by J. H. Rowen,[4] the mechanism of dielectric loss in most ferrites is well understood, and it is now relatively easy to manufacture a ferrite today with practically any desired composition whose dielectric Q is of the order of 1,000. Even materials such as Ferroxcube III-C, which for a long time were considered to be extremely lossy, can now be manufactured so that the dielectric loss at microwave frequencies is vanishingly small. To be sure, the low-frequency properties (initial permeability) are usually degraded by slightly altering the composition and method of preparation, but the fundamental magnetic properties which are of importance at microwave frequencies are change negligibly. Thus one need not be greatly concerned with dielectric losses in ferrites except for high power applications.

In addition, all magnetic losses were ignored, except those associated with the imaginary part of the permeability which describes the phenomenon of ferromagnetic resonance. Now it is well known from the experimental and theoretical work of G. T. Rado[5] at the Naval

[4] J. H. Rowen, "Ferrites in Microwave Applications," *Bell System Tech. Jour.*, vol. 32, p. 1333; 1953.

[5] G. T. Rado, "Ferromagnetism at very high frequencies," *Phys. Rev.* vol. 80, p. 273; 1950.

Research Laboratories and J. Smit and D. Polder[6] at Philips Laboratories that demagnetized pieces of ferrite have magnetic losses which are not accounted for by the above theory. In the first place, highly sintered ferrites, in which domain wall motion takes place in weak fields, show large losses due to wall motion at frequencies which correspond to the natural resonant frequency of the walls. It is to be expected that the natural frequency of domain walls could vary over extremely wide frequency limits, depending upon the physical and chemical structure of the particular ferrite. Adequate experimental evidence is not available today to verify this expectation, since the only clear-cut experiments seem to have been performed by G. T. Rado and J. K. Galt.[7] In a sample of magnesium ferrite, Dr. Rado showed that a domain wall resonance occurred at approximately 50 mc and that appreciable losses due to this mechanism extended up to frequencies of at least 500 mc and possibly higher. In addition, demagnetized ferrites have domain rotation losses which can extend up through X-band frequencies. These losses are really ferromagnetic resonance absorption, with the natural frequency being determined by effective internal magnetic fields. In general, these losses can extend from

$$\omega_{min} = \gamma H_a \tag{28}$$

to

$$\omega_{max} = \gamma(H_a + 4\pi M_s), \tag{29}$$

where

H_a = effective anisotropy field (approximately equal to

$$\frac{2K_1}{M_s}),$$

M_s = saturation magnetic moment of material.

Rado's data showing these two resonance regions is shown in Fig. 3.

Obviously, when a sample of ferrite is completely magnetized by a dc magnetic field, it cannot show domain wall losses for small superposed rf magnetic fields; and hence one might assume that losses which occur in demagnetized ferrites are of little concern to microwave applications where the ferrite is used in a magnetized condition. However, such is not always the case, because these "zero field" losses are not completely removed until dc magnetic fields are applied, which are appreciably greater than that which is necessary to essentially magnetize the material. If now one attempts to build the isolator described above to operate at 130 mc, then the optimum front to back ratio occurs in an applied magnetic field of 45 oersteds, and this probably is not a large enough field to completely erase domain wall losses. Since all zero field losses are reciprocal (i.e.,
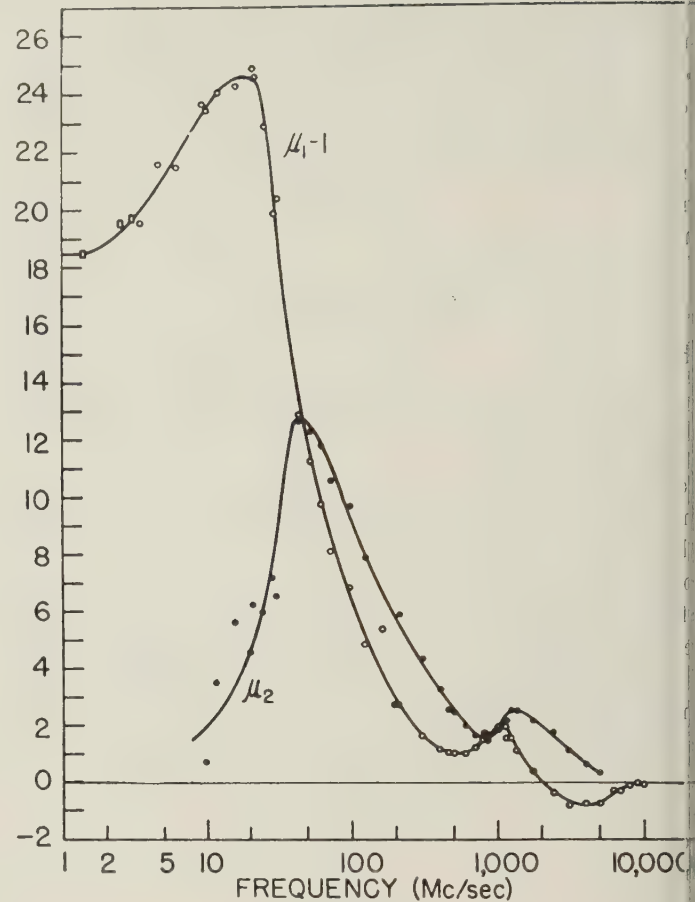
Fig. 3

independent of the sense of the circular polarization), the presence of such losses deteriorate the performance of all microwave ferrite devices.

L. G. Van Uitert, J. P. Shafer, and the author[8] have shown that if one is operating sufficiently far above the minimum frequency given in (28) but still below the maximum frequency given in (29), the zero field domain rotation losses also disappear as the ferrite is magnetized. In addition, if one is operating at sufficiently high frequencies (above 2000 mc), it is usually possible to reduce the maximum frequency at which these losses occur by reducing the saturation moment of the material according to (29). However, for frequencies below 1,000 mc, it is evident that this approach becomes useless, not only because the saturation moment is becoming vanishingly small (and hence the material is becoming less ferromagnetic), but according to (29), ω_{max} cannot be reduced to less than γH_a regardless of the saturation moment.

It thus becomes evident that a complete understanding of the operation of microwave ferrite devices below 1,000 mc must wait until a more complete understanding of domain wall and domain rotation losses in demag-

[6] J. Smit and D. Polder, "Resonance phenomenon in ferrites," *Rev. Mod. Phys.*, vol. 25, p. 89; 1953.
[7] J. K. Galt, "Motion of domain walls in ferrite crystals," *Rev. Mod. Phys.*, vol. 25, p. 93; 1953.

[8] L. G. Van Uitert, J. P. Shafer, and C. L. Hogan, "Low-loss ferrites for applications at 4,000 mc/sec," *Jour. Appl. Phys.*, vol. 25, p. 925; 1954.

netized ferrites is obtained. What is especially needed is an understanding of how a ferrite can be manufactured so that these losses can be completely eliminated or moved to frequency regions which are of no concern.

The entire theory of nonreciprocal devices has so far been developed by neglecting anisotropy fields. Such an approximation is valid for frequencies high enough so that

$$f \gg 2.8\left(\frac{2K_1}{M_s}\right), \qquad (30)$$

where

K_1 = first-order anisotropy constant in ergs/cc,
f = frequency in mc.

However, for operation below 1,000 mc, the first-order anisotropy constant of many ferrites is so high that this relation is not valid, and hence the theory as developed above cannot be applied. Thus, if we again consider an infinite polycrystalline ferrite medium (so as to neglect demagnetizing fields), the effect of crystal anisotropy is simply that of broadening the resonance absorption line provided (30) is observed. However, if the frequency of operation is sufficiently less than that given in (30), the problem is complicated, since some crystallites which are aligned with their "easy" direction of magnetization along the Z-axis (direction of applied dc field) always have a resonant frequency greater than the applied frequency, regardless of the applied magnetic field. Crystallites which are aligned with the "hard" direction of magnetization along the Z-axis actually have their resonant frequency reduced by the applied magnetic field, until the applied magnetic field reaches a value

$$H_{app} = \frac{2K_1}{M_s}. \qquad (31)$$

Under ideal conditions, the resonant frequency of such crystallites is zero at this point, and it then increases linearly with the applied field. Thus, if the frequency of operation is less than

$$f = 2.8\left(\frac{2K_1}{M_s}\right) \qquad (32)$$

the entire sample cannot be at resonance at any given field strength; and as the frequency is gradually reduced below that given above, the ferromagnetic resonance line gradually disappears. When the frequency is so low that

$$f \ll 2.8\left(\frac{2K_1}{M}\right), \qquad (33)$$

then essentially all crystallites in the sample have resonant frequencies greater than the applied frequency, regardless of the applied field. The results obtained from such a material would be similar to a material used at higher frequencies but biased with a dc magnetic field which places the material well above resonance. In reference to Figs. 1 and 2, it would correspond to that material
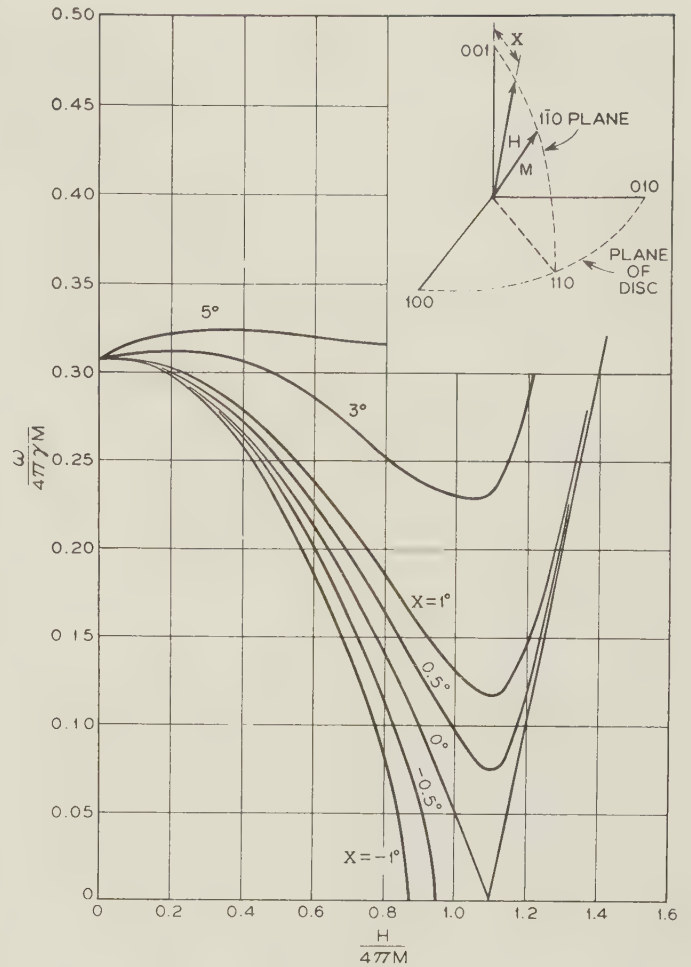
Fig. 4

being biased with a magnetic field well above resonance (i.e., 5,000 oersteds). Under this condition, the effective permeability which the ferrite presents to the positive circularly polarized wave is not sensibly different from that which it presents to the negative circularly polarized wave, and the ferrite thus loses its anti-reciprocal property.

There are only two methods apparent for overcoming this difficulty for low-frequency operation. One is to use single crystals aligned with the hard direction of magnetization along the Z-axis and to make use of the fact that the resonant frequency can be reduced by the application of the applied field. H. Suhl has calculated the resonant frequency of such a crystal which is cut in the 001 plane. These curves are shown in Fig. 4. This data was calculated for a material in which

$$\frac{2K_1}{M} = 240 \text{ oersteds}, \qquad 4\pi M = 3600 \text{ gauss}.$$

It will be noticed from this data that if the applied magnetic field is accurately aligned along the 001 direction, then the resonant frequency can be reduced sufficiently far so that microwave devices can be built to operate easily at their ultimate low frequency limit of

100 to 200 mc, regardless of the crystal anisotropy. However, for the case which Suhl calculated, if the applied magnetic field is tilted by 30 minutes from the 001 direction, the minimum resonant frequency which the disk has is about 750 mc̃. This occurs in an applied field of approximately 3,800 eresteds. Thus it appears that the geometrical perfection required of the ferrite crystal and associated waveguide parts is so critical that it would be impractical to attempt to build devices which operate below 500 mc by using this technique.

The other method of overcoming the difficulty which arises from using "high anisotropy" ferrites is to prepare ferrites with vanishingly small anisotropies. These can now be prepared in materials with relatively low Curie temperatures, and there is considerable hope that this might be accomplished in materials with relatively high Curie temperatures by making mixed ferrites of cobalt ferrite with some of the other ferrite materials.

Low Frequency Limitation of Nonreciprocal Phase Shifter

A nonreciprocal phase shifter must operate in a region as far removed from the ferromagnetic resonance line as is possible in order to minimize the losses associated with this phenomenon. The ideal phase shifter would have zero attenuation, but a nonreciprocal phase shifter using ferrites must have a finite attenuation, since its operation depends upon the fact that the ferrite has an absorption line for a positive circularly polarized wave. To obtain the low frequency limit of an anti-reciprocal phase shifter we need only calculate the differential phase shift per decibel loss and then find how low a frequency we can go to and still have acceptable operation.

In referring to Figs. 1 and 2, it is obvious that a nonreciprocal phase shifter can be operated in two regions. The first and most satisfactory region is in weak dc fields, where the ferrite is just magnetized but the applied field is still too weak to bring the ferrite into the vicinity of the absorption line. As will be shown below, this becomes increasingly difficult as one goes to lower and lower frequencies. When this is so, then it becomes evident from Figs. 1 and 2 that in magnetic fields which are considerably larger than that required for resonance (i.e., 4,000 oersteds for the particular case illustrated in Figs. 1 and 2, the total attenuation can again be brought close to zero; but the ferrite still displays a permeability for the positive circular component which is substantially different from that for the negative circularly polarized component, and hence anti-reciprocal action is still possible. R. H. Fox and B. Lax[9] of Lincoln Laboratory have already shown how operation "above resonance" can be used to extend the low frequency limit of certain geometries. We shall compute the two cases separately.

[9] R. H. Fox and B. Lax, Mass. Inst. Tech., Lincoln Laboratory, Group 37, Tech. Rept. No. 65 ⋯⋯⋯

Again, as in the case of a ferrite isolator, there are several different microwave geometries that could be investigated, and they would all lead to different low frequency limits; but again we shall find it more illuminating to examine the intrinsic properties of the material first, and from this information we shall find it relatively easy to examine particular geometries in detail. The most fundamental geometry is identical to that discussed with respect to the ferrite isolator and is shown in Fig. 5. Again in this case, a longitudinal magnetic field is applied to the ferrite, but it is biased into a region to give differential phase shift rather than differential attenuation. The wave, while propagating through the ferrite in one direction, is a positive circularly polarized wave, and while propagating in the opposite direction, is a negative polarized wave.
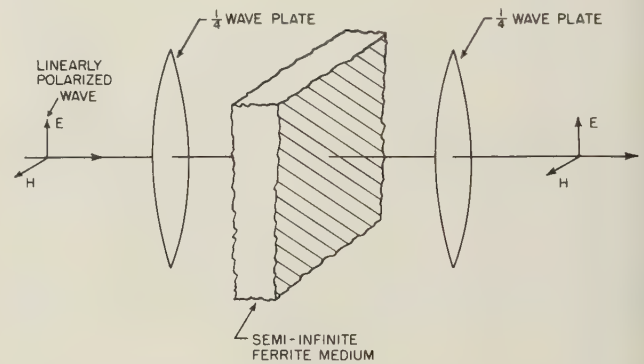
Fig. 5—Magnetic spectrum of solid ferramic A in the demagnetized slate. Data were taken on coaxial lines (circles) and on a Gir. 821A bridge (rectangles). The value of $(\mu::::-1)$ is 18.6.

Now the differential phase shift per unit path length through the ferrite is given by

$$\theta = [\beta_- - \beta_+], \qquad (34)$$

where

$\beta_\pm$ = propagation constant for positive and negative circular component.

The propagation constants are given by (6) above, and provided (12) is valid, these can be written as

$$\beta_\pm = \omega \sqrt{\epsilon \mu_o \mu_\pm'}. \qquad (35)$$

Hence (34) can be written as

$$\theta = \omega \sqrt{\epsilon \mu_o} [\sqrt{\mu_-'} - \sqrt{\mu_+'}] \qquad (36)$$

In calculating the insertion loss per unit length, we shall neglect the attenuation of the negative component and shall calculate the loss per unit length for the positive component, since under ideal conditions this component suffers a greater loss. The attenuation of the positive component per unit length is given by (13). Thus the differential phase shift per decibel loss is given by

$$\frac{\theta}{L_+} = \frac{\sqrt{\mu_-'\mu_+'} - \mu_+'}{4.4\,\mu_+''}. \qquad (37)$$

Substituting from (1) to (3), this formula becomes

$$\frac{\theta}{L_+} = \frac{\sqrt{\left(1 + \dfrac{4\pi M\gamma}{\gamma H - \omega}\right)\left(1 + \dfrac{4\pi M\gamma}{\gamma H + \omega}\right) - \left(1 + \dfrac{4\pi M\gamma}{\gamma H - \omega}\right)}}{4.4 \dfrac{4\pi\lambda\omega}{(\gamma H - \omega)^2}}, \tag{38}$$

which is valid for all field strengths, provided (12) is valid.

In vanishingly small applied magnetic fields (low field side of resonance), (38) becomes

$$\frac{\theta}{L_+} = \frac{\left[\sqrt{1 - \left(\dfrac{4\pi M\gamma}{\omega}\right)^2} - 1 + \dfrac{4\pi M\gamma}{\omega}\right]\omega}{17.6\pi\lambda}. \tag{39}$$

Now if

$$\frac{4\pi M\gamma}{\omega} \ll 1, \tag{40}$$

(39) takes the particularly simple form

$$\frac{\theta}{L_+} = \frac{2H_r}{4.4\Delta H}, \tag{41}$$

where (23) has been used to replace λ by the resonance line width in oersteds.

Eq. (41) can easily be shown to hold, even when the relation given in (40) is reversed. However, when $\frac{4\pi M\gamma}{\omega} \simeq 1$, no such simple relation can be derived; but the actual relation predicts the same low-frequency limit, and hence this relation can be considered as valid under all conditions, provided one is operating on the low field side of resonance.

Now to build a gyrator, a differential phase shift of π radians is required, and if it is intended to accomplish this with an insertion loss of less than one-half decibel, then

$$\frac{2H_r}{4.4\Delta H} \geqq 2\pi, \tag{42}$$

$$H_r \geqq 13.7(\Delta H). \tag{43}$$

For a line width of 50 oersteds,

$$H_r \geqq 685 \text{ oersteds.} \tag{44}$$

This means that a gyrator cannot be built with an insertion loss less than one-half decibel that will operate at frequencies below 1,900 mc, unless some geometry can be found which is more favorable than that of an infinite plane wave in a semi-infinite medium, or unless we operate at field strengths greater than that required for resonance.

Actually, a 45-degree Faraday rotator requires only half the differential phase shift of a gyrator, and yet circulators and other nonreciprocal elements can be constructed from this element. Thus this particular geometry gives a factor of 2, and hence the low-fre-

quency limit of a circulator can be placed at approximately 1,000 mc from infinite plane wave analysis for operation in weak magnetic fields.

Now for operation in magnetic fields which are large enough, so that

$$\gamma H_z \gg \omega, \tag{45}$$

(36) can be written as

$$\theta = -\frac{\omega^2\sqrt{\epsilon\mu_o}}{\gamma H_z}\sqrt{1 + \frac{4\pi M}{H}}, \tag{46}$$

and (13) can be written as

$$L_+ = \frac{17.4\omega^2\sqrt{\mu_o\epsilon}\,(\pi\lambda)}{\gamma^2 H_z^2\sqrt{1 + \dfrac{4\pi M}{H}}}. \tag{47}$$

Thus the differential phase shift per unit loss becomes

$$\frac{|\theta|}{L_+} = \frac{H_r}{2.2\Delta H}\left(1 + \frac{H_z}{4\pi M}\right)^{10}. \tag{48}$$

Eq. (48) shows that the differential phase shift per unit loss can be increased without limit by increasing the applied magnetic field above resonance. The only limiting factor is the size of the element that can be used practically, since the differential phase shift per unit path length decreases as the magnetic field is increased. Eq. (48) predicts that a differential phase shift of $\pi/2$ can be obtained with an insertion loss of one-half decibel in a device operating at 500 mc, provided

$$H_z \simeq 4\pi M_s. \tag{49}$$

Conclusion

The above analysis is intended to indicate many of the problems which exist and must be solved in order to decrease the present low frequency limit of microwave ferrite devices. It certainly is not all-inclusive, and each particular waveguide geometry offers its own particular problem which usually arises because of the effect of demagnetizing factors upon resonant frequency. In addition, guided wave modes usually do not consist of pure positive and negative circularly polarized fields, and hence further complications are introduced in applying the results of the above anaysis to wave guide devices. Nevertheless, it is felt that this analysis should give engineers an appreciation of the intrinsic properties of the materials themselves, and only very clever designing will allow one to produce elements whose properties are better than the infinite plane wave devices discussed above.

Acknowledgment

[10] Actually (48) is valid only when $H_z \ll 4\pi M$ so that the low frequency limit above resonance is not sensibly different from that below resonance.

Some Topics in the Microwave Application of Gyrotropic Media

A. A. van TRIER†

Summary—The Faraday effect of plane and guided waves is reviewed in Sections I and II. Section III deals with a cavity technique for measuring Faraday rotations in a circular waveguide with a coaxial ferrite pencil. In Section IV some experimental results are discussed, including the evaluation of the permeability tensor components, the relation between Faraday rotation and pencil radius, and ferromagnetic resonance in circularly polarized waves. The problem of the rectangular gyrotropic waveguide is taken up in Section V. A simple method of successive approximations is described and applied to the case of the square waveguide.

I. Introduction

NONRECIPROCAL anisotropic, or simply gyrotropic, media have been investigated lately by many authors, both from a theoretical and from an experimental point of view. The applicability of nonreciprocal ferrites, especially at microwave frequencies, is now well established, and the number of nonreciprocal components incorporating ferrites is rapidly increasing. One of these elements, the unidirectional waveguide, has found widespread use, whereas some others have proved their usefulness and are likely to find other applications. In this paper some topics will be discussed, scattered more or less, which refer to work done in the Netherlands in this field of research.

Gyrotropic media will be defined by the relations

$$
\begin{aligned}
B_x &= \mu_1 H_x - j\mu_2 H_y \,, \\
B_y &= j\mu_2 H_x + \mu_1 H_y \,, \\
B_z &= \mu_o H_z \,,
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
D_x &= \epsilon_1 E_x - j\epsilon_2 E_y \,, \\
D_y &= j\epsilon_2 E_x + \epsilon_1 E_y \,, \\
D_z &= \epsilon_3 E_z \,.
\end{aligned}
\tag{2}
$$

Eq. (1) applies to ferrites magnetized in the z direction, (2) to solid dielectrics ($\epsilon_2 \ll \epsilon_1$) or hf gas discharges in a constant magnetic field H_o along the z axis. It is easy to see that for circularly polarized plane waves the medium has a scalar permeability and a scalar dielectric constant. For, if we substitute

$$
\begin{aligned}
H_y &= \pm jH_x, & H_z &= 0, \\
E_y &= \pm jE_x, & E_z &= 0
\end{aligned}
$$

in (1) and (2), we obtain

$$
B_x = (\mu_1 \pm \mu_2)H_x, \quad B_y = (\mu_1 \pm \mu_2)H_y, \quad B_z = H_z = 0,
$$
$$
D_x = (\epsilon_1 \pm \epsilon_2)E_x, \quad D_y = (\epsilon_1 \pm \epsilon_2)E_y, \quad D_z = E_z = 0,
\tag{3}
$$

or

$$
\overline{B} = (\mu_1 \pm \mu_2)\overline{H}, \quad \overline{D} = (\epsilon_1 \pm \epsilon_2)\overline{E}.
\tag{4}
$$

The upper signs correspond to counterclockwise circularly polarized waves, the lower signs to clockwise

circularly polarized waves. The propagation constants of these waves are

$$
\beta_\pm = \omega[(\mu_1 \pm \mu_2)(\epsilon_1 + \epsilon_2)]^{1/2} \text{ meter}^{-1}.
\tag{5}
$$

As is well-known, superposition of a clockwise and a counterclockwise circularly polarized wave of equal amplitudes, but with $\beta_+ \neq \beta_-$, results in a linearly polarized wave that shows a Faraday rotation

$$
\Theta = (\beta_+ - \beta_-)/2 \text{ radians per meter.}
\tag{6}
$$

The nonreciprocity of the medium finds expression here in the fact that the direction of rotation is independent of the direction in which the wave traverses the medium.

These results are easy to obtain, but complications arise if we consider gyrotropic media in waveguides. We shall restrict ourselves from now on to media that are gyromagnetic only.

II. Guided Waves in Gyromagnetic Media

Although some important problems in the theory of guided wave propagation in gyrotropic media have been solved, many questions still remain to be answered. The problem of the circular waveguide containing a coaxial ferrite bar magnetized along the axis has received most attention, and for this configuration the characteristic equation determining the eigenvalues of the propagation constant is known.[1-4] This equation, however, is a complicated one, from which numerical information can be obtained only at the cost of a considerable amount of computing. For very thin ferrite pencils perturbation theory leads to a result.

The case of the rectangular waveguide with transversely magnetized ferrites can be dealt with rigorously for TE_{on} waves and the like,[1,4] but, except for the completely filled waveguide, it is very laborious to obtain numerical results.

The situation is least satisfactory for the case of a rectangular or square waveguide with longitudinal magnetization. Here only methods of successive approximation lead to some results.

† Philips Research Laboratories, N. V. Philips' Gloeilampen-fabrieken, Eindhoven-Netherlands.

[1] H. Suhl and L. R. Walker, "Topics in guided-wave propagation through gyromagnetic media," *Phys. Rev.*, vol. 86, p. 122; 1952; *Bell Sys. Tech. Jour.*, vol. 33, pp. 579–659; May, 1954; pp. 939–986; July, 1954; pp. 1133–1194; September, 1954.
[2] H. J. Gamo, "The Faraday rotation of waves in a circular waveguide," *Jour. Phys. Soc.*, Japan, vol. 8, pp. 176–182; March, 1953.
[3] M. L. Kales, "Modes in waveguides containing ferrites," *Jour. Appl. Phys.*, vol. 24, pp. 604–608; May, 1953.
[4] A. A. van Trier, "Guided electromagnetic waves in anisotropic media," *Appl. Sci. Res.*, vol. 3, Section B, pp. 305–371; December, 1953.

On the other hand it may be pointed out that a qualitative picture of what is going on can often be obtained by very simple physical arguments, which lead to the same conclusions as the rigorous mathematical analysis. As an example, let us consider a completely filled circular waveguide. In an isotropic circular waveguide we are more or less free to consider solutions with $\cos n\Phi$, $\sin n\Phi$, or solutions with $\exp(\pm jn\Phi) = \cos n\Phi \pm j \sin n\Phi$ as fundamental. These can be called "linearly" and "circularly" polarized modes respectively. In view of what we have seen in Section I for plane waves in gyrotropic media it is unlikely that there will be modes which are able to maintain a simple $\cos n\Phi$ – $\sin n\Phi$ distribution while traveling along the waveguide. So we expect that in gyrotropic media the fundamental solutions have Φ dependence $\exp(\pm jn\Phi)$, and moreover $\beta_+ \neq \beta_-$. Superposition of these solutions results in a wave with alternately $\cos(n\Phi)$ and $\sin (n\Phi)$ distribution; i.e., a linearly polarized mode with "Faraday rotation."

TE_{11}

$\cos \Phi \qquad \pm j \sin \Phi$

TE_{21}

$\cos 2\Phi \qquad \pm j \sin 2\Phi$

TM_{11}

$\cos \Phi \qquad \pm j \sin \Phi$
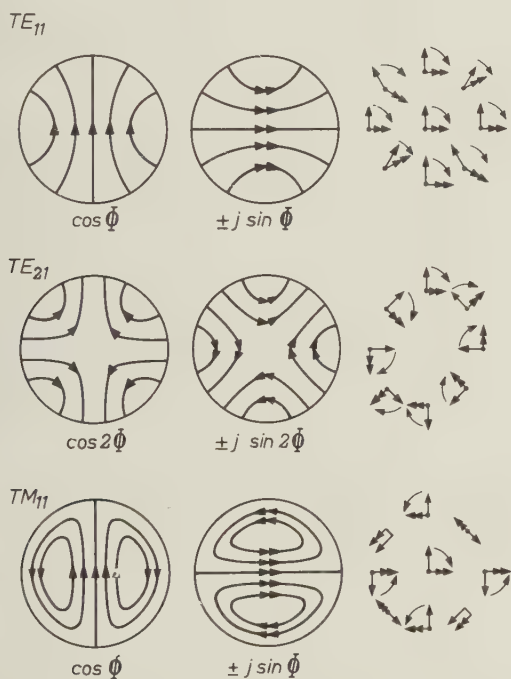
Fig. 1—Circularly polarized isotropic modes. The transverse field pattern of various modes is shown as well as the direction of rotation of the field. Note that the electric field is drawn for the TE modes, and the magnetic field for the TM mode.

By studying the transverse field distributions of the various circularly polarized isotropic modes (Fig. 1), we can easily understand why some modes show a large Faraday rotation, whereas others show only a small effect. If the transverse field of the isotropic modes with $\exp (jn\Phi)$ and $\exp (-jn\Phi)$, respectively, has at all points the same direction of rotation, circular or elliptical, then it follows from (3) that on making the medium gyrotropic there will be an increase of μ at all points for one mode and a decrease of μ for the other mode. This results in a large difference between the effective permeabilities for the two circularly polarized modes, and, therefore, in a large Faraday rotation. If, however, the direction of rotation of the field is not the same at all points of the cross section, an area of increasing μ will counteract an area of decreasing μ, resulting in a small difference $\beta_+ - \beta_-$. According to this intuitive reasoning the Faraday effect will be large for the TE_{11} mode and the TE_{21} mode, but small for the TM_{11} mode. Indeed these results are confirmed by the rigorous mathematical analysis.

In practical applications completely filled circular waveguides cannot be used for several reasons. In the case of large gyrotropy there is a considerable difference between the transverse field distributions and the propagation constants β_+, β_- of the two circularly polarized modes, and this makes it very difficult to match both modes simultaneously. Another, more fundamental, complication is the fact that higher modes are likely to be excited. This is made possible by the high dielectric constant of ferrites ($\epsilon \approx 10$), and the occurrence of higher modes is even stimulated by the gyrotropic components (mode conversion). These difficulties can be avoided by using thin coaxial ferrite pencils. The matching problem is then easily solved by relatively short tapers and if the diameter is chosen sufficiently small there is no possibility of mode conversion.
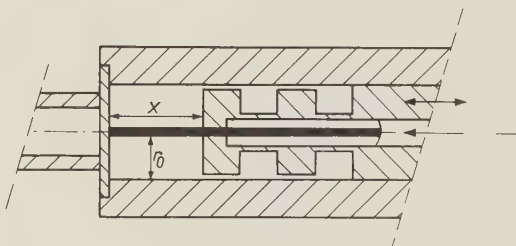
Fig. 2—Cavity with coaxial ferrite pencil for use at 24000 mc. The cavity radius $r_o = 0.5$ cm.

III. AN INDIRECT METHOD OF MEASURING FARADAY ROTATION

The Faraday rotation of guided waves in a circular waveguide with a coaxial ferrite pencil can be measured directly, as has been done by Roberts,[5] Hogan[6] and many others. It is possible also, however, to measure the wavelengths $\lambda_{g\pm}$ of the two circularly polarized waves. From these we calculate $\beta_\pm = 2\pi/\lambda_{g\pm}$ and obtain finally $\Theta = (\beta_+ - \beta_-)/2$. The wavelength measurements are most easily done by means of a cavity. To obtain accurate results a cavity of special form is required (Fig. 2).

[5] F. F. Roberts, "A note on the ferromagnetic Faraday effect at centimetre wavelengths," *Jour. Phys. Radium*, vol. 12, pp. 305–307; March, 1951.

[6] C. L. Hogan, "The ferromagnetic Faraday effect at microwave frequencies and its applications; the microwave gyrator," *Bell. Sys. Tech. Jour.*, vol. 31, pp. 1–31; January, 1952; *Rev. Mod. Phys.*, vol. 25, pp. 253–263; January, 1953.

This TE_{11n} circular cavity can be tuned by means of a movable piston sliding freely over the ferrite pencil which is kept in position by external means. The cavity terminates a waveguide and the reflection coefficient R is measured with a directional coupler DC (Fig. 3). The microwave source, a $2K33$ Raytheon reflex klystron, is connected to the system via a unidirectional waveguide. The cavity is coupled with the waveguide by means of two circular coupling holes located symmetrically with respect to the E plane of the rectangular waveguide. Resonances of the cavity are observed as "dips" in the reflection R.
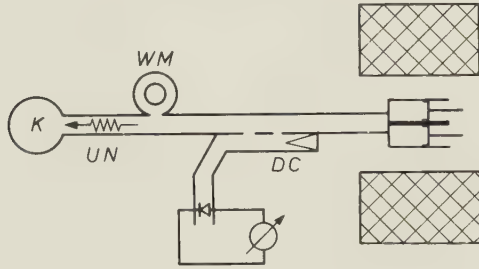
Fig. 3—Block diagram of experimental set-up.

The hole in the piston sets an upper limit to the ferrite diameters that can be used: it must form a waveguide below cutoff to prevent loading of the cavity. As long as the ferrite is not magnetized it is isotropic and the cavity resonates in the degenerate TE_{11} mode when $x = n\lambda_g/2$. As soon as the magnetic field is applied the degeneracy is removed ($\lambda_{g+} \neq \lambda_{g-}$) and each of the original resonances splits into two resonances at $x = n\lambda_{g\pm}/2$.

If the allowed displacement of the piston is large enough to cover subsequent resonances, the end effects of the cavity are completely eliminated and a very accurate measurement of β_+ and β_- is possible. Fig. 4 shows some experimental data obtained with a K-band cavity of 1 cm inner diameter. In this cavity Ferroxcube pencils with diameters ranging from 0.7 to 3.0 mm have been investigated. The maximum error in the Faraday rotation Θ is of the order of 0.07 degrees per centimeter.

It may be added that a resonance splitting similar to the one described here has been observed by Beljers[7] in a TE_{11} cavity containing a flat disk instead of a pencil.

IV. DISCUSSION OF EXPERIMENTAL RESULTS

With the perturbation theory it can be shown that the propagation constants of the circularly polarized waves in circular waveguides containing thin coaxial ferrite pencils are given by the expression

$$\beta_\pm = \beta_o + \Delta\beta_\pm = \beta_o + r_1^2\left[C_1\frac{\mu_1 \pm \mu_2 - \mu_o}{\mu_1 \pm \mu_2 + \mu_o} + C_2\frac{\epsilon - \epsilon_o}{\epsilon + \epsilon_o}\right].$$

(7)

[7] H. G. Beljers, "Faraday effect in magnetic materials with travelling and standing waves," *Philips Res. Reps.*, vol. 9, pp. 131–139; April, 1954.

Here β_o is the propagation constant in the empty waveguide, r_1 is the radius of the ferrite pencil, and the constants $C_{1,2}$ are known functions of the waveguide dimensions and of the frequency. It is seen that the corrections are proportional to r_1^2.
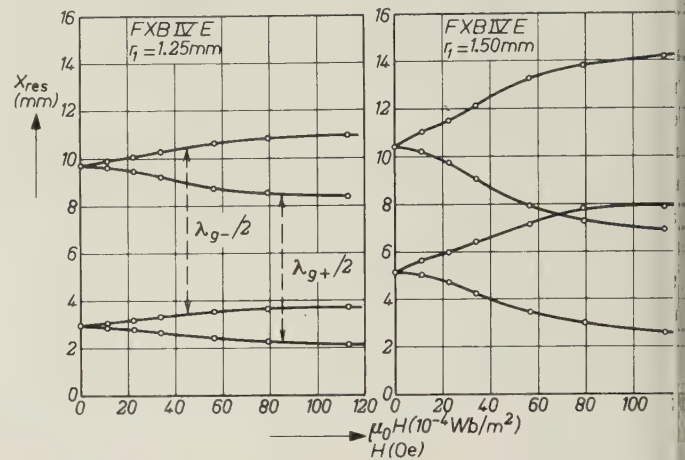
Fig. 4—Resonance splitting vs applied magnetic field for pencils of Ferroxcube IV E with radii of 1.25 and 1.50 mm respectively.

A. Evaluation of μ_1 and μ_2

It is evident that in principle (7) can be used to obtain experimental values of μ_1 and μ_2 once ϵ is known. It turns out, unfortunately, that for the thinnest ferrite pencils that can be used (7) is not yet valid. It is possible, however, to measure $\Delta\beta_\pm$ for various radii r_1 and to extrapolate the values of $\Delta\beta_\pm/r_1^2$ towards $r_1 = 0$. Some results obtained for specimens of Ferroxcube IV have been collected in Table I. In these experiments the ferrites were magnetized by a field of about 110 Oe which is sufficient to nearly saturate them.

TABLE I

Ferroxcube	Theory		Experiment	
	μ_1/μ_o	μ_2/μ_o	μ_1/μ_o	μ_2/μ_o
IV A	1	0.285	1.01 ± 0.03	0.31 ± 0.03
IV B	1	0.47	1.10 ± 0.06	0.54 ± 0.05
IV C	1	0.43	1.05 ± 0.05	0.49 ± 0.05
IV D	1	0.30	1.01 ± 0.03	0.31 ± 0.03
IV E	1	0.20	0.98 ± 0.02	0.20 ± 0.01

The experimental data can be compared with theoretical values obtained from the Polder[8] relations for saturated ferromagnetics

$$\mu_1 = \mu_o + \frac{\gamma^2 M_o H_o}{\gamma^2 H_o^2 - \omega^2}, \qquad \mu_2 = \frac{\omega\gamma M_o}{\gamma^2 H_o^2 - \omega^2}.$$ (8)

Here γ is the negative gyromagnetic ratio, H_o the constant magnetic field by which the medium is saturated, and M_o the saturation magnetization.

[8] D. Polder, "On the theory of ferromagnetic resonance," *Phil. Mag.*, vol. 40, pp. 99–115; January, 1949.

in our case $-\gamma H_o \ll \omega$, and $-\gamma M_o \leqslant \omega$; therefore

$$\mu_1 \approx \mu_o, \qquad \mu_2 \approx -\frac{\gamma M_o}{\omega}. \qquad (9)$$

The agreement between the experimental and theoretical values in Table I is seen to be satisfactory.

Ferromagnetic Resonance

One of the circularly polarized modes must show ferromagnetic resonance absorption for a critical value of the applied magnetic field. This means that one of the μ must show a singularity. It is evident from (8) that μ_1 and μ_2 resonate for $\omega = -\gamma H_o$. In that case, however, β_- and $\Delta\beta_-$ remain finite, corresponding to the fact that the resonance frequency is shifted to higher values due to the demagnetizing fields. To find the modified resonance condition we consider the expressions $\mu_1 \pm \mu_2 + \mu_o$. Substituting (8), we obtain

$$\mu_1 + \mu_2 + \mu_o = 2\mu_o + \frac{\gamma M_o}{\gamma H_o - \omega}, \qquad (10a)$$

$$\mu_1 - \mu_2 + \mu_o = 2\mu_o + \frac{\gamma M_o}{\gamma H_o + \omega}. \qquad (10b)$$

Clearly $\mu_1 + \mu_2 + \mu_o$ is always positive, but $\mu_1 - \mu_2 + \mu_o$ can become zero. So we find as resonance conditions

$$\omega = -\gamma\left[H_o + \frac{M_o}{2\mu_o}\right], \quad \text{or} \quad H_{o_{res}} = -\frac{\omega}{\gamma} - \frac{M}{2\mu_o}. \quad (11)$$
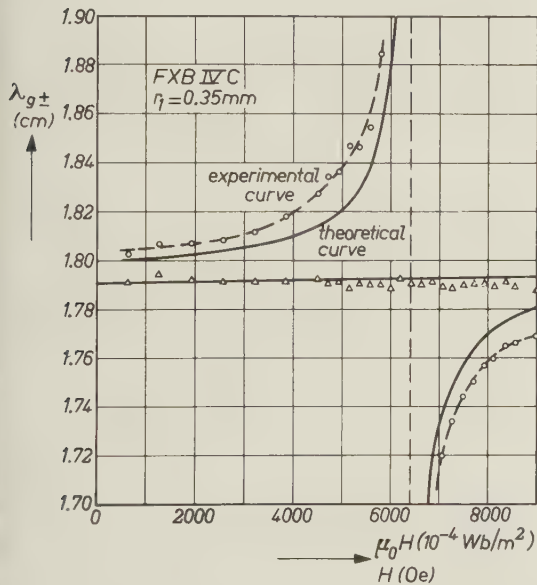
Fig 5—Ferromagnetic resonance in a Ferroxcube IV C pencil radius $r_1 = 0.35$ mm.

It will be noted that these expressions are identical with those derived by Kittel for an infinitely long thin ferrite cylinder in a homogeneous hf magnetic field. Fig. 5 shows the results of a resonance experiment done at the Kamerlingh Onnes Laboratory at Leiden with an air-coil in which fields up to 10^4 Oe could be produced. The graph shows the experimental points as well as the theoretical curves obtained from (7) and (8).

C. Large diameter pencils

We have seen already that (7) applies to ferrite pencils with very small diameters only. In practical applications we shall often use ferrite pencils whose radius is 0.1 to 0.3 times the waveguide radius. As it is difficult to obtain theoretical results for this case we determined the relation between Faraday rotation and pencil diameter experimentally. It was rather surprising to find that in this range of radii the Faraday rotation follows more or less the empirical law

$$\log \Theta = c_1 + c_2 r_1, \qquad (12)$$

as is shown in Fig. 6. This result is confirmed by data published by Fox, Miller and Weiss.[9]
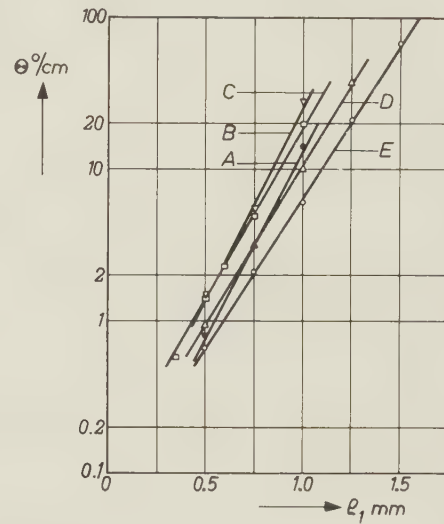
Fig. 6—Faraday rotation vs pencil radius for average radius pencils.

V. RECTANGULAR AND SQUARE WAVEGUIDES

As we have seen in Section II which dealt with circular waveguides, gyrotropy causes coupling between normally degenerate modes with $\cos n\Phi$ and $\sin n\Phi$ distributions. Using the same type of intuitive reasoning as applied to the circular waveguide problem, we must expect in square waveguides a similar coupling between the TE_{01} mode and the TE_{10} mode with transverse electric fields

$$E_x = E_x{}^{01} \sin \frac{\pi y}{a}, \qquad E_y = 0;$$

respectively

$$E_x = 0, \qquad E_y = E_y{}^{10} \sin \frac{\pi x}{a}.$$

There is Faraday effect if the field distribution along the traveling wave is alternately TE_{01} and TE_{10}.

A satisfactory rigorous mathematical treatment of the rectangular waveguide problem has not yet been

[9] A. G. Fox, S. E. Miller, and M. T. Weiss, "Behavior and applications of ferrites in the microwave region," *Bell Sys. Tech. Jour.*, vol. 34, pp. 5–105; January, 1955.

given. Some conclusions can be drawn from Schelkunoff's generalized mode theory,[10] but this method is very laborious indeed. A very simple method of successive approximations which will be outlined below for the completely filled square waveguide, has been used by De Vries and the author. We first substitute (3) in Maxwell's equations and obtain

$$\nabla \times \overline{E} = -j\omega\overline{B} = -j\omega//\mu//\overline{H},$$
$$\nabla x \overline{H} = j\omega\overline{D} = j\omega\epsilon\overline{E}. \qquad 13)$$

(The time-factor is supposed to be exp $j\omega t$).

The usual procedure which consists in eliminating the transverse field components from (13) leads to two simultaneous equations for the longitudinal components of the form

$$\nabla_t^2 E_z + aE_z + bH_z = 0, \qquad (14)$$
$$\nabla_t^2 H_z + cE_z + dH_z = 0,$$

where $\nabla_t^2 \equiv \nabla^2 - \dfrac{\partial^2}{\partial z^2}$, and a, b, c, and d are constants

depending on ω, μ and ϵ. As is well known these equations can be solved for the circular but not for the rectangular boundary-value problem. Rather than following the above approach we now eliminate H_x, H_y, H_z and

$$E_z = \frac{1}{j\beta}\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y}\right)$$

from (13) and obtain the equations

$$0_{xx}E_x + 0_{xy}E_y = -j\frac{\mu_2}{\mu_1}\left[\frac{\partial^2 E_x}{\partial x \partial y} + \left(\frac{\partial^2}{\partial y^2} - \beta^2\right)E_y\right],$$
$$\qquad (15)$$
$$0_{yx}E_x + 0_{yy}E_y = j\frac{\mu_2}{\mu_1}\left[\left(\frac{\partial^2}{\partial x^2} - \beta^2\right)E_x + \frac{\partial^2 E_y}{\partial x \partial y}\right],$$

where

$$0_{xx} \equiv \omega^2\epsilon\mu_1\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) + \frac{\partial^2}{\partial x^2} + \frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right)\frac{\partial^2}{\partial y^2} - \beta^2,$$

$$0_{xy} = 0_{yx} \equiv -\left[\frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) - 1\right], \qquad (16)$$

$$0_{yy} \equiv \omega^2\epsilon\mu_1\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) + \frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right)\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \beta^2.$$

We must try to find solutions of (15) that satisfy the boundary conditions (cf. Fig. 7)

$$E_x = 0 \quad \text{at} \quad y = 0, b;$$
$$E_y = 0 \quad \text{at} \quad x = 0, a; \qquad (17)$$

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = 0 \quad \text{at} \quad x = 0, a; \quad y = 0, b.$$

[10] S. A. Schelkunoff, "Generalized telegraphist's equations for waveguides," Bell. Sys. Tech. Jour., vol. 31, pp. 784–801; July, 1952.

The general solution of (15) which satisfies (17) is the form

$$E_x = \sum_{mn} E_x^{mn} \cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b},$$
$$E_y = \sum_{mn} E_y^{mn} \sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b}. \qquad (1$$

To obtain the zeroth-order approximations we put the right-hand side of (15) equal to zero. As is easily verified we then obtain the sets of TE_{mn} and TM_{mn} modes. One of these is selected and substituted in the right-hand side of (15), and the first-order approximation is found.
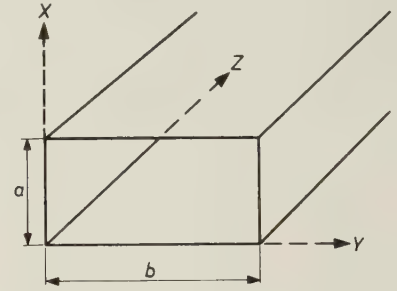
Fig. 7—Rectangular waveguide with co-ordinate system.

This process can be contained indefinitely. Evidently nothing can be said about the convergence of the process, but it seems at least justified to assume this convergence if $\mu_2/\mu_1 \ll 1$. We shall now consider the case of the square waveguide in some detail.

As zeroth-order approximation we take a combination of a TE_{01} and a TE_{10} mode

$$E_x = E_x^{01}\sin\frac{\pi y}{a}; \quad E_y = E_y^{10}\sin\frac{\pi x}{a}. \qquad (19$$

These modes have the same propagation constant

$$\beta_0^2 = \omega^2\epsilon\mu_1\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) - \frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right)\frac{\pi^2}{a^2}. \qquad (19$$

Next we substitute (19) in the right-hand side, and (18) in the left-hand side of (15), using the abbreviations

$$\omega^2\epsilon\mu_1\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) - \frac{m^2\pi^2}{a^2} - \frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right)\frac{n^2\pi^2}{a^2}$$
$$- \beta^2 = M_{mn},$$

$$\left[\frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) - 1\right]\frac{m\pi}{a}\frac{n\pi}{a} = P_{mn}, \qquad (2$$

$$\omega^2\epsilon\mu_1\left(1 - \frac{\mu_2^2}{\mu_1^2}\right) - \frac{\mu_1}{\mu_0}\left(1 - \frac{\mu_2^2}{\mu_1^2}\right)\frac{m^2\pi^2}{a^2}$$
$$- \frac{n^2\pi^2}{a^2} - \beta^2 = N_{mn}.$$

The result is

$$\sum_{mn}(M_{mn}E_x^{mn} + P_{mn}E_y^{mn})\cos\frac{m\pi x}{a}\sin\frac{n\pi y}{a}$$
$$= j\frac{\mu_2}{\mu_1}\beta_0^2 E_y^{10}\sin\frac{\pi x}{a}, \qquad (2$$

$$\sum_{m,n} (P_{mn}E_x{}^{mn} + N_{mn}E_y{}^{mn}) \sin \frac{m\pi x}{a} \cos \frac{n\pi y}{a}$$

$$= -j \frac{\mu_2}{\mu_1} \beta_0{}^2 E_x{}^{01} \sin \frac{\pi y}{a}.$$

Now it is known that, if $0 < x < a$

$$\sin \frac{u\pi x}{a} = \sum_v S_u^v \cos \frac{v\pi x}{a}, \qquad u + v \text{ odd},$$

$$\cos \frac{u\pi x}{a} = \sum_v C_u^v \sin \frac{v\pi x}{a}, \qquad u + v \text{ odd}; \quad (22a)$$

with

$$S_u^v = \frac{4}{\pi} \frac{u}{u^2 - v^2}; \qquad S_u^0 = \frac{2}{\pi}\frac{1}{u}; \qquad C_u^v = -\frac{4}{\pi}\frac{v}{u^2 - v^2}. \quad (22b)$$

Substituting (22) in (21) we obtain

$$\sum_{m,n} (M_{mn}E_x{}^{mn} + P_{mn}E_y{}^{mn}) \cos \frac{m\pi x}{a} \sin \frac{n\pi y}{a}$$

$$= j\frac{\mu_2}{\mu_1}\beta_0{}^2 E_y{}^{10} \sum_p S_1{}^{2p} \cos \frac{2p\pi x}{a} \sum_q C_0{}^{2q+1} \sin \frac{(2q+1)\pi y}{a},$$

$$\sum_{m,n} (P_{mn}E_x{}^{mn} + N_{mn}E_y{}^{mn}) \sin \frac{m\pi x}{a} \cos \frac{n\pi y}{a} \qquad (23)$$

$$= -j\frac{\mu_2}{\mu_1}\beta^2 E_x{}^{01} \sum_q C_0{}^{2q+1} \sin \frac{(2q+1)\pi x}{a} \sum_p S_1{}^{2p} \cos \frac{2p\pi y}{a}.$$

Equating coefficients of equal functions in (23), we find that

$$M_{2p,2q+1}E_x{}^{2p,2q+1} + P_{2p,2q+1}E_y{}^{2p,2q+1}$$

$$= j\frac{\mu_2}{\mu_1}\beta^2 E_y{}^{10}S_1{}^{2p}C_0{}^{2q+1},$$

$$P_{2p,2q+1}E_x{}^{2p,2q+1} + N_{2p,2q+1}E_y{}^{2p,2q+1} = 0,$$

$$M_{2q+1,2p}E_x{}^{2q+1,2p} + P_{2q+1,2p}E_y{}^{2q+1,2p} = 0, \qquad (24)$$

$$P_{2q+1,2p}E_x{}^{2q+2,2p} + N_{2q+1,2p}E_x{}^{2q+1,2p}$$

$$= -j\frac{\mu_2}{\mu_1}\beta_0{}^2 E_x{}^{01}C_0{}^{2q+1}S_1{}^{2p};$$

and moreover, that

$$E_x{}^{2p,2q} = E_y{}^{2p,2q} = E_x{}^{2p+1,2q+1} = E_y{}^{2p+1,2q+1} = 0.$$

$$(24^1)$$

In particular we have for $p = q = 0$,

$$M_{01}E_{x0}{}^1 = (\beta_0{}^2 - \beta^2)E_x{}^{01} = j\frac{\mu_2}{\mu_1}\beta_0{}^2 E_y{}^{10}S_1{}^0 C_0{}^1, \qquad (25)$$

$$N_{10}E_y{}^{10} = (\beta_0{}^2 - \beta^2)E_y{}^{10} = -j\frac{\mu_2}{\mu_1}\beta_0{}^2 E_x{}^{01}C_0{}^1 S_1{}^0.$$

From (25) it follows immediately that

$$E_y{}^{10} = \pm jE_x{}^{01}, \qquad (26)$$

and

$$\beta_\pm{}^2 = \beta_0{}^2\left[1 \pm \frac{\mu_2}{\mu_1}\frac{8}{\pi^2}\right]. \qquad (27)$$

These formulas confirm the results of our intuitive reasoning and describe the coupling between the TE_{01} and TE_{10} modes. Using the fact that for square waveguides $M_{2p,2q+1} = N_{2q+1,2p}$, and $P_{2p,2q+1} = P_{2q+1,2p}$ we obtain for the amplitudes of the higher harmonics

$$E_x{}^{2p,2q+1} = \mp jE_y{}^{2q+1,2p}$$

$$= \mp \frac{\mu_2}{\mu_1}\frac{N_{2p,2q+1}\beta_0{}^2 E_x{}^{01}S_1{}^{2p}C_0{}^{2q+1}}{M_{2p,2q+1}N_{2p,2q+1} - P^2{}_{2p,2q+1}},$$

$$E_y{}^{2p,2q+1} = \mp jE_x{}^{2q+1,2p} \qquad (28)$$

$$= \pm \frac{\mu_2}{\mu_1}\frac{P_{2p,2q+1}\beta_0{}^2 E_x{}^{01}S_1{}^{2p}C_0{}^{2q+1}}{M_{2p,2q+1}N_{2p,2q+1} - P^2{}_{2p,2q+1}}.$$

It may be remarked that the same analysis has been applied to rectangular waveguides. It is found that in a rectangular waveguide no mode coupling occurs in general, and that the perturbation of β is of the order of $(\mu_2/\mu_1)^2$.

The Seismic Pulse, an Example of Wave Propagation in a Doubly Refracting Medium

C. L. PEKERIS†

Summary—An exact and closed solution is given for the motion produced on the surface of a uniform elastic half-space by the sudden application of a concentrated pressure-pulse at the surface. The time variation of the applied stress is taken as the Heaviside unit function, and its concentration at the origin is such that the integral of the force over the surface is finite. This problem gives an instructive illustration of wave propagation in a doubly refracting medium, since both shear waves and compressional waves are excited, and they travel with different speeds. There is, in addition, the Rayleigh surface wave. For a medium in which the elastic constants λ and μ are equal, the vertical component of displacement w_o at the surface is given by:

$$w_o = 0, \qquad \tau < \frac{1}{\sqrt{3}},$$

$$w_o = -\frac{Z}{\pi\mu r}\left\{\frac{3}{16} - \frac{\sqrt{3}}{32\sqrt{\tau^2 - \frac{1}{4}}} - \frac{\sqrt{5 + 3\sqrt{3}}}{32\sqrt{\frac{3}{4} + \frac{\sqrt{3}}{4} - \tau^2}}\right.$$
$$\left. + \frac{\sqrt{3\sqrt{3} - 5}}{32\sqrt{\tau^2 + \frac{\sqrt{3}}{4} - \frac{3}{4}}}\right\} \qquad \frac{1}{\sqrt{3}} < \tau < 1,$$

$$w_o = -\frac{Z}{\pi\mu r}\left\{\frac{3}{8} - \frac{\sqrt{5 + 3\sqrt{3}}}{16\sqrt{\frac{3}{4} + \frac{\sqrt{3}}{4} - \tau^2}}\right\}, \qquad 1 < \tau < \frac{1}{2}\sqrt{3 + \sqrt{3}},$$

$$w_o = -\frac{Z}{\pi\mu r}\frac{3}{8}, \qquad \tau > \frac{1}{2}\sqrt{3 + \sqrt{3}},$$

where $\tau = (ct/r)$, c-shear wave velocity, and $-Z$ is the surface integral of the applied stress.

The horizontal component of displacement is obtained similarly in terms of elliptic functions. A discussion is given of the various features of the waves.

It is pointed out that in the case of a buried source, an observer on the surface will, under certain circumstances, receive a wave which travels to the surface as an S wave along the ray of total reflection, and from there along the surface as a *diffracted P wave*. An exact expression is given for this diffracted wave.

The question of the suitability of automatic computing machines for the solution of pulse propagation problems is also discussed.

Introduction

THE COMPLEXITY of the problem of propagation of waves generated impulsively in doubly refracting media, is due not so much to the duplicity of the waves, as to the fact that when one type of wave reaches a boundary it is generally partially converted into a wave of the other type. One can conceive of, and sometimes construct, impulsive sources which will generate one type of wave only, and such a pure type of wave will actually propagate from the source until such time as it reaches the first boundary. After that time the original type of wave will, generally, be accompanied by the second type of wave, traveling in a different direction. For the understanding of the nature of these intricate phenomena it is sometimes helpful to study other doubly refracting media, where particular features of the wave problem may lend themselves more readily to experimental or analytical in-

† Department of Applied Mathematics, The Weizmann Institute, Rehovot, Israel.

vestigation. It is with this in mind that I wish to bring to your attention some recent results in the theory of seismic wave propagation.

Consider a uniform elastic half-space characterized by the elastic constants λ and μ, and the density ρ. Such a medium can propagate two types of elastic waves: transverse equivoluminal shear waves S, and longitudinal compressional waves P. The speeds c and c_p of the S and P waves respectively are given by

$$c^2 = (\mu/\rho) \qquad c_p^2 = (\lambda + 2\mu)/p_1 = 3c^2, \qquad (1)$$

where we have assumed the elastic constants λ and μ to be equal. The separation into two types of waves can be exhibited by the fact that it is possible to express the vertical component of displacement w and the horizontal component q (in cylindrical co-ordinates) by the relations

$$q = \varphi_r + \chi_{rz}, \qquad w = \varphi_z + \chi_{zz} - k^2\chi, \qquad (2)$$

where the potential φ satisfies the equation for compressional waves, and the potential χ the equation for S waves:

$$\nabla^2\varphi - (p^2/c_p^2)\varphi = 0, \quad \nabla^2\chi - k^2\chi = 0, \quad k = p/c. \quad (3)$$

Here p denotes the operator $(\partial/\partial t)$, and the subscripts denote partial differentiation. The stresses also involve both φ and χ:

$$p_{zz} = \lambda(p/c_p)^2\varphi + 2\mu(\varphi_{zz} + \chi_{zzz} - k^2\chi_z), \qquad (4)$$

$$p_{rz} = u(\partial/\partial r)(2\varphi_z + 2\chi_{zz} - k^2\chi), \qquad (5)$$

so that when one applies a concentrated normal stress p_{zz} at the source, both compressional φ waves and shear χ waves are excited.

We shall consider the case when the difference in stress Δp_{zz} across the horizontal plane becomes infinite at the source in such a manner that

$$2\pi\int_0^\infty \Delta p_{zz}r\,dr = Z, \qquad (6)$$

Z being a (negative) constant. The time-variation of the applied stress at the source will be assumed to be given by the Heaviside unit function $H(t)$.

The Ray Picture for the Buried Pulse

Our discussion will be concerned mainly with the case when the source is situated on the surface. We shall, however, interpolate in this section some remarks on certain physical features of the motion of the surface produced by a *buried source*. Let the source be situated at a depth H below the surface, as shown in Fig. 1. Consider the shear wave S excited by the source. When this wave reaches the surface, it is reflected partially as the S' wave, and partially as the P' wave, shown in Fig. 1(a), for the case $r < H/\sqrt{2}$. P' deviates more from

the vertical than does S'. In Fig. 1(b), $r = H/\sqrt{2}$, and the derived P' wave is directed along the surface. For $r > H/\sqrt{2}$, the incident S wave is totally reflected, with S' having the same amplitude as S, but a different phase. There is no reflected P' wave. However, the exact solution of this boundary value problem[1] yields the result that before the time of arrival of the direct S wave, a point on the surface, such as B in Fig. 1(c), receives a *diffracted* wave which traveled via the path OAB, traversing the leg OA with the speed c of shear waves, and the leg AB with the speed c_p of compressional waves. On the basis of the ray theory, no such wave would be expected, since all of the incident energy in the case of Fig. 1(b) is carried away by the reflected S' wave. If one plots the time of arrival t of this diffracted wave against the epicentral distance r, the slope of the straight line so obtained is of course $(1/c_p)$, and the
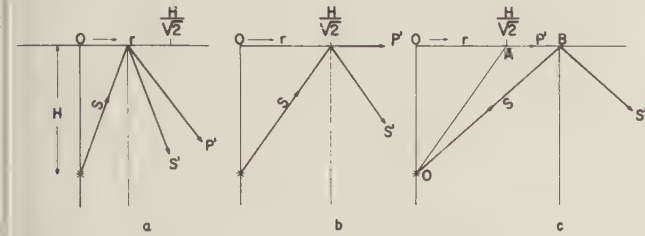
Fig. 1—Ray diagram of the reflection of an SV wave at a free surface. S denotes incident SV wave, S' the reflected shear wave, and P' the derived compressional wave. In (a), $r < H/\sqrt{2}$. In (b) S is totally reflected, and a diffracted P' travels along the surface. In (c), ($r > H/\sqrt{2}$) the point B receives, before the arrival of the direct S wave, a diffracted P' wave which traveled along OA with the shear velocity c, and along AB with the compressional velocity $\sqrt{3}c$. The diagram was drawn under the assumption that the elastic constants λ and μ are equal.

intercept of this line with the axis of t is $\sqrt{2/3}(H/c)$, from which the depth of the source H can be determined.

Mathematically this appears in the form of the following operational result entering into the expression for the displacement[1]

$$\int_0^\infty J_o(krx)f(x)x\,exp(-kH\sqrt{x^2+1})\,dx \to 0,$$
$$\tau < \tau^*, \quad (7)$$

$$\to -(2/\pi R)\int_{1/\sqrt{3}}^{\epsilon(\tau)} f(iv)v[h^2 + \tau^2 - v^2$$
$$- 2h\tau\sqrt{1-v^2}]^{-1/2}\,dv, \quad \tau^* < \tau < 1, \quad (8)$$

$$\to -(2/\pi R)\,I\int_{1/\sqrt{3}}^{v(\tau)} f(iv)v[h^2 + \tau^2 - v^2$$
$$- 2h\tau\sqrt{1-v^2}]^{-1/2}\,dv, \quad \tau > 1, \quad (9)$$

with

$$= (p/c), \quad R = \sqrt{r^2 + H^2}, \quad h = (H/R),$$
$$= (ct/R), \quad \tau^* = [\sqrt{2/3}\,h + (1/\sqrt{3})\sqrt{1-h^2}], \quad (10)$$

$$(\tau) = (\tau\sqrt{1-h^2} - h\sqrt{1-\tau^2}),$$
$$v(\tau) = (\tau\sqrt{1-h^2} - ih\sqrt{\tau^2-1}). \quad (11)$$

[1] C. L. Pekeris, "The seismic buried pulse," *Proc. Nat. Acad. Sci.*, vol. 41, pp. 629–631; 1955.

Here the arrow denotes operational interpretation, I signifies the imaginary part, and $f(x)$ is an even function of x having branch-points at $x = \pm(i/\sqrt{3})$ and $\pm i$. These branch-points are associated with the existence of the compressional waves and shear waves respectively. τ^* corresponds to the time of arrival of the diffracted wave OAB of Fig. 1(c), and τ to the time of arrival of the S wave OB. Eq. (8) holds only for $r > H/\sqrt{2}$, while for $r < H/\sqrt{2}$ [Fig. 1(a)] it vanishes.
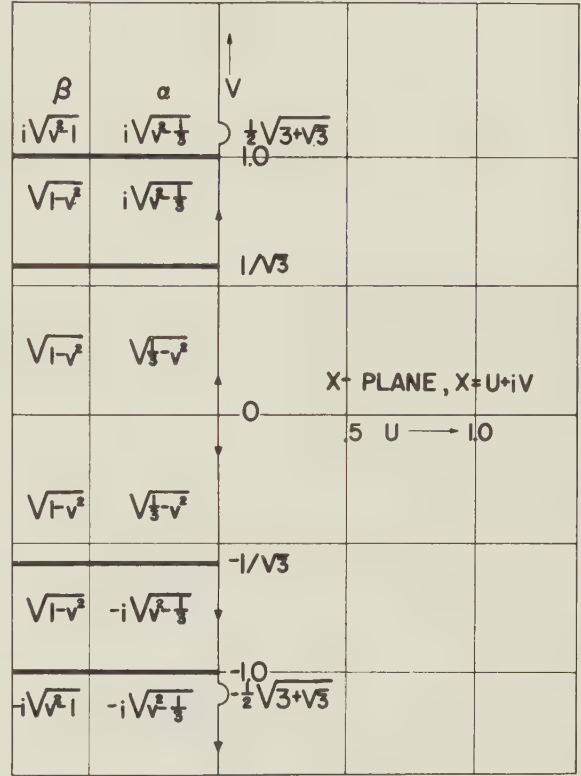
Fig. 2—The complex x-plane is cut up by branch-lines (thick) at the branch points $v = \pm(1/\sqrt{3})$ and $v = \pm 1.0$. The poles are situated at $v = \pm(1/2)\sqrt{3 + \sqrt{3}}$. α and β denote the values of $\sqrt{1/3 + x^2}$ and of $\sqrt{1 + x^2}$ on the various portions of the v-axis.

The Surface Pulse

When the point-source is situated at the surface and is integrable in the sense of (6), the operational representation of the vertical component of displacement at the surface $w(p)$ is given by[2]

$$w(p) = (Zk/2\pi\mu)\int_o^\infty J_o(krx)m(x)x\,dx, \quad (12)$$

with

$$m(x) = \alpha/[(2x^2+1)^2 - 4x^2\alpha\beta], \quad (13)$$

$$\alpha = \sqrt{x^2 + \tfrac{1}{3}}, \quad \beta = \sqrt{x^2 + 1}, \quad k = (p/c). \quad (14)$$

By cutting up the complex x-plane in the manner shown in Fig. 2, one makes the function $m(x)$ regular to the right of the imaginary axis, enabling one to apply to the integral in (12) the Bateman-Pekeris theorem,[3] to

[2] C. L. Pekeris, "The seismic surface pulse," *Proc. Nat. Acad. Sci.*, vol. 41, pp. 469–480; 1955.
[3] H. Bateman and C. L. Pekeris, "Transmission of light from a point source in a medium bounded by diffusely reflecting parallel plane surfaces," *Jour. Opt. Soc. Amer.*, vol. 35, pp. 651–657; 1945.

the effect that

$$\int_0^\infty J_o(krx)xm(x)\,dx$$

$$= -(2/\pi)I \int_0^\infty K_o(krv)vm(iv)\,dv. \qquad (15)$$

With the known interpretation of

$$kK_o(krv) \to 0, \qquad v > \tau = (ct/R),$$

$$\to (1/r)(\tau^2 - v^2)^{-1/2} \qquad v < \tau, \qquad (16)$$

we thus get

$$\int_0^\infty J_o(krx)xm(x)\,dx \to 0 \to 1/\sqrt{3} \qquad \tau < 1\sqrt{3}, \quad (17)$$

$$\to -(2/\pi r)IP \int_{1/\sqrt{3}}^{\pi} vm(iv)(\tau^2 - v^2)^{-1/2}\,dv,$$

$$\tau > 1/\sqrt{3}, \qquad (18)$$

where P denotes principal value.

By rationalization and partial fraction decomposition one can evaluate the integral in (18) in closed form,[4] as follows:

$$w(\tau) = 0,\, 1/\sqrt{3} \qquad \tau < 1\sqrt{3}, \qquad (19)$$

$$w(\tau) = -(Z/32\pi\mu r)\left\{6 - \frac{\sqrt{3}}{\sqrt{\tau^2 - \frac{1}{4}}}\right.$$

$$\left. - \frac{\sqrt{3\sqrt{3}+5}}{\sqrt{\frac{3}{4}+\frac{\sqrt{3}}{4}-\tau^2}} + \frac{\sqrt{3\sqrt{3}-5}}{\sqrt{\tau^2 + \frac{\sqrt{3}}{4}-\frac{3}{4}}}\right\}, \qquad (20)$$

$$\frac{1}{\sqrt{3}} < \tau < 1,$$

$$w(\tau) = -(Z/16\pi\mu r)\left\{6 - \frac{\sqrt{3\sqrt{3}+5}}{\sqrt{\frac{3}{4}+\frac{\sqrt{3}}{4}-\tau^2}}\right\}, \qquad (21)$$

$$1 < \tau < \gamma = (1/2)\sqrt{3+\sqrt{3}},$$

$$w(\tau) = -(3Z/8\pi\mu r), \qquad \tau > \gamma. \qquad (22)$$

The operational representation of the horizontal component of displacement $q(p)$ in the case of a surface-pulse of the form $H(t)$, and subject to condition (6), is given by

$$q(p) = (Z/2\pi\mu)(\partial/\partial r)\int_0^\infty J_o(krx)n(x)x\,dx, \qquad (23)$$

$$n(x) = [(2x^2 + 1) - 2\alpha\beta]/[(2x^2 + 1)^2 - 4x^2\alpha\beta]. \qquad (24)$$

By following the same methods outlined above, one arrives at the result

$$q(\tau) = 0, \qquad \tau < 1/\sqrt{3}, \qquad (25)$$

$$q(\tau) = (\sqrt{3/2}\, Z/16\pi^2\mu r)\{6K(k) - 18\Pi(8k^2, k)$$

$$+ (6 - 4\sqrt{3})\Pi[-(12\sqrt{3} - 20)k^2, k]$$

$$+ (6 + 4\sqrt{3})\Pi[(12\sqrt{3} + 20)k^2, k]\},$$

$$1/\sqrt{3} < \tau < 1, \qquad (26)$$

[4] C. L. Pekeris, *Proc. Nat. Acad. Sci.*, vol. 26, pp. 433–436; 1940. For details of the derivation, see reference 2.

$$q(\tau) = (\sqrt{3/2}\tau\kappa Z/16\pi^2\mu r)\{6K(\kappa) - 18\Pi(8, \kappa)$$

$$+ (6 - 4\sqrt{3})\Pi[-(12\sqrt{3} - 20), \kappa]$$

$$+ (6 + 4\sqrt{3})\Pi[(12\sqrt{3} + 20), \kappa]\},$$

$$1 < \tau < \gamma = (1/2)\sqrt{3+\sqrt{3}}, \qquad (27)$$

$$q(\tau) = (\sqrt{3/2}\tau\kappa Z/16\pi^2\mu r)\{6K(\kappa) - 18\Pi(8, \kappa)$$

$$+ (6 - 4\sqrt{3})\Pi[-(12\sqrt{3} - 20), \kappa]$$

$$+ (6 + 4\sqrt{3})\Pi[(12\sqrt{3} + 20), \kappa]\} + \frac{Z\tau}{8\pi\mu r\sqrt{\tau^2 - \gamma^2}},$$

$$\tau > \gamma, \qquad (28)$$

with

$$k^2 = (1/2)(3\tau^2 - 1), \qquad \kappa = (1/k), \qquad (29)$$

$$K(k) = \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2\sin^2\theta}},$$

$$\Pi(n,k) = \int_0^{\pi/2} \frac{d\theta}{(1 + n\sin^2\theta)\sqrt{1 - k^2\sin^2\theta}}. \qquad (30)$$

The function $\Pi(n, k)$ can be expressed in terms of incomplete elliptic integrals, for which tables are available.
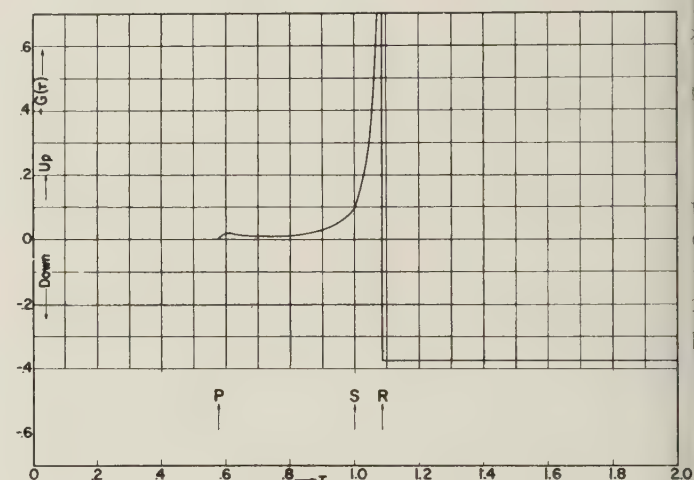
Fig. 3—Vertical displacement at the surface $w(t)$ due to the application of a Heaviside unit pressure-pulse $H(t)$ at the surface. $w(t) = -(Z/\pi\mu r)G(\tau)$, $\tau = (c_s t/r)$. Z (negative) is the surface integral of the applied pressure. $w(t)$ was computed from (19)–(22). P denotes arrival time of the compressional wave, S of the shear wave, and R of the Rayleigh wave.

DISCUSSION OF RESULTS

Fig. 3 shows the vertical displacement $w(\tau)$, as computed from (19)–(22). The initial motion, due to the application of a downward pressure at the source, is upwards, as shown at the epoch P, when the compressional wave reaches the point. The arrival of the shear wave S is marked only by a change of slope in $w(\tau)$. R marks the arrival of the Rayleigh *surface* wave. Just prior to this time, $w(\tau)$ becomes infinite, with an amplitude varying with range r like $1/\sqrt{r}$, which is characteristic of surface wave propagation. The steady state displacement is downwards, and is reached right after the arrival of the Rayleigh wave.

Fig. 4 shows the horizontal displacement $q(\tau)$ com-
[pu]ted from (25)–(28). The initial displacement is out-
[wa]rds, but the final displacement is inwards. The arrival
[of] the S wave is characterized again only by a change
[of] slope, which is even less marked here.
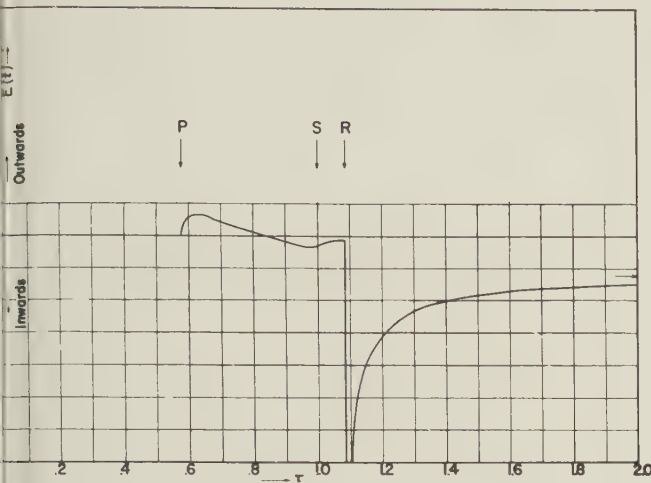
[Fig.] 4—Horizontal displacement at the surface $q(t)$ due to the
[ap]plication of a Heaviside unit pressure-pulse $H(t)$ at the surface.
[$q(t)$] $= -(Z/\pi\mu r)E(\tau)$, $\tau = (c_s t/r)$, Z (negative) is the surface
[int]egral of the applied pressure. $q(t)$ was computed from (25)–(28).
[P,] S and R denote times of arrival of compressional, shear, and
[Ra]yleigh waves respectively.

[ON] THE FEASIBILITY OF USING ELECTRONIC COMPUTERS
FOR THE SOLUTION OF IMPULSIVE WAVE
PROPAGATION PROBLEMS

[O]ne of the aims in solving the seismic buried source
[pr]oblem for a uniform elastic half-space was to obtain
[an] exact solution with which to test the accuracy of a
[dir]ect numerical method of solution using automatic
[co]mputing machinery. Were this approach feasible,

then one could cope with the more interesting case of a
layered half-space. This problem is now under investiga-
tion, and it is hoped that the results will prove to be more
encouraging than preliminary considerations[4] led us to
expect. For, if we write

$$w(\tau) = -(Z/\pi\mu r)F(\tau), \qquad (31)$$

then it can be shown that (12) is equivalent to the inte-
gral equation

$$m(x) = 2\int_0^\infty F(\tau)\tau(x^2 + \tau^2)^{-3/2}\,d\tau, \qquad (32)$$

which can be verified by substituting for $F(\tau)$ from
(19)–(22).

It is found that the trial solution $F_1(\tau)$ given by

$$F_1(\tau) = 0, \qquad\qquad \tau < .719574516, \quad (33)$$

$$F_1(\tau) = -.0612441866;$$
$$.719574516 < \tau < 1.167093607, \quad (34)$$

$$F_1(\tau) = (3/8), \qquad\qquad \tau > 1.167093607, \quad (35)$$

when substituted into (32) yields a function $m_1(x)$ such
that the maximum deviation of $[m_1(x)/m(x)]$ from 1 is
1.00086, around $x = 1$, and that this ratio is closer to
unity for all other values of x. One might be inclined to
expect that $[F_1(\tau)/F(\tau)]$ is equally close to unity.
Actually, $F_1(\tau)$ is a very bad solution from the point of
view of the seismologist, for it yields a P-wave velocity
which is wrong by 25 per cent; it lacks completely the
S phase as for a liquid medium, and the Rayleigh wave
velocity is wrong by 7 per cent. This R phase has fur-
thermore a variation with distance of $1/r$, rather than
the correct variation of $1/\sqrt{r}$. The source of our diffi-
culty lies in the fact that we are interested primarily
in minute details of only isolated portions of the func-
tion, rather than in its over-all behavior.

On the Electromagnetic Characterization of Ferromagnetic Media: Permeability Tensors and Spin Wave Equations

GEORGE T. RADO†

Summary—Various constitutive equations applicable to ferromagnetic and ferrimagnetic media are discussed systematically, the emphasis being on a formulation and analysis of the underlying assumptions. A distinction is made between the "ordinary" (Maxwellian) and certain "average" field vectors. The latter are useful in the presence of domain structure; they include appropriately defined spatial averages, $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$, of the time-dependent components of the ordinary $\vec{B}$ and $\vec{H}$, respectively. In cases where $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ are connected by a "point relation", the general form of Polder's permeability tensor is extended to nonsaturated media; the special tensors due to Polder, the writer, and Wangsness, are then reviewed. In cases where $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ are not so connected, the "exchange effect" and the "spin wave equation" are discussed. Following Ament and Rado, three consequences of this equation are treated: the new boundary conditions, and the triple refraction and "equivalent isotropic permeability" in metals.

I. Introduction

IT IS well-known that under certain conditions the theory of electromagnetic wave propagation in ferromagnetic media is rather complicated. Such conditions occur, for example, in the case of microwaves propagating in a ferromagnetic medium which is contained in a waveguide and possesses a static magnetization. The mathematical difficulties and results relevant to this case have been reviewed in an excellent article published by van Trier.[1] His paper, as well as the more recent ones by Suhl and Walker,[2] discusses the original theoretical contributions of Gamo, Kales, Suhl and Walker, and van Trier.

In contrast to the work of the authors mentioned, the present paper is not primarily concerned with the mathematical problems of solving Maxwell's equations and satisfying the corresponding boundary conditions. Instead, the present paper deals mainly with the physical problems associated with the constitutive equations in ferromagnetic media. The nature of these equations, *i.e.*, the manner in which ferromagnetic media are characterized electromagnetically, is evidently the root of the mathematical difficulties referred to above. Furthermore, the limitations of the constitutive equations may cause ambiguities in the experimentally measured electromagnetic properties of a ferromagnetic medium. It seems appropriate, therefore, to discuss these equations in a systematic and unified manner.

The only essentially new contribution made in this paper is an attempt to establish and justify a general form for the permeability tensor in nonsaturated ferromagnetic media, i.e., in media characterized by a domain structure. Since this attempt is purposely not based on any special physical model, the presentation of the new theory is tantamount to a somewhat lengthy formulation and analysis of the underlying assumptions. These assumptions, treated in Section II, include a distinction between the ordinary (Maxwellian) field vectors and certain "average" field vectors recently introduced by the writer.[3]

The quantities $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ will be defined later on being certain spatial averages of the time-dependent components of $\vec{B}$ and $\vec{H}$, respectively. In Section III those equations are treated that connect $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ what will be referred to as a "point relation." Such relation is exemplified by the permeability tensor. Its general form is derived in Section IIIA, and its various special forms, originally derived by the writer,[3] Polder, and Wangsness,[5] are reviewed (together with some background material) in Section IIIB.

There is one equation, treated in Section IV, th connects $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ by other than a point relatic This so-called "spin wave equation" and the underlyi exchange effect are briefly reviewed in Section IV. Certain consequences of this equation, namely t triple refraction, new boundary conditions, and "equiv lent isotropic permeability" in ferromagnetic meta are discussed in Section IVB on the basis of Ament ar Rado's recent work.[6]

Finally, Appendixes I and II contain some remar on the conductivity tensor and on the calculation energy losses, respectively.

II. Some Assumptions and Definitions

Several assumptions and definitions are formulat and analyzed below. They will be denoted by lower ca letters [e.g., (a)], and will be understood to apply throug out this paper unless an explicit statement is made to t contrary.

A. Ferromagnetism vs Ferrimagnetism

(a) It is assumed that there is no qualitative differen between the electromagnetic behavior of ferromagnet

† Naval Research Laboratory, Washington, D. C.
[1] A. A. Th. M. van Trier, "Guided electromagnetic waves in anisotropic media," *Appl. Sci. Res.*, sec. B, vol. 3, pp. 305–371; April, 1953.
[2] H. Suhl and L. R. Walker, "Topics in guided wave propagation through gyromagnetic media I, II, III," *Bell. Sys. Tech. Jour.*, vol. 33, pp. 579–659, 939–986, 1133–1194; May, July, and September, 1954.

[3] G. T. Rado, "Theory of the microwave permeability tens and Faraday effect in nonsaturated ferromagnetic material *Phys. Rev.*, vol. 89, p. 529; January, 1953.
[4] D. Polder, "On the theory of ferromagnetic resonance *Phil. Mag.*, vol. 40, pp. 99–115; January, 1949.
[5] R. K. Wangsness, "Susceptibility tensor and the Faraday effe in ferrimagnetics," *Phys. Rev.*, vol. 95, pp. 339–345; July, 1954.
[6] W. S. Ament and G. T. Rado, "Electromagnetic effects of sp wave resonance in ferromagnetic metals," *Phys. Rev.*, vol. 97, p 1558–1566; March, 1955.

and ferrimagnetic media. For the sake of brevity, both will be referred to as being ferromagnetic.

A brief discussion of (a) is given in Section IIIB3. It will be seen there that only in one exceptional kind of medium, that which is characterized by "compensation points," do the typically ferrimagnetic properties of the medium have a qualitative (rather than merely quantitative) effect on its electromagnetic behavior.

B. The Ordinary Field Vectors and their Interrelations

(b) The "ordinary field vectors" are defined to be the field vectors $\vec{H}, \vec{B}, \vec{E}, \vec{D}$ occurring in the Maxwell theory. The meaning of these (macroscopic) vectors is briefly discussed below.

The four field vectors and the current density $\vec{J}$ must satisfy the two Maxwell equations

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c} \frac{\partial \vec{B}}{\partial t}, \tag{1}$$

$$\vec{\nabla} \times \vec{H} = \frac{4\pi}{c} \vec{J} + \frac{1}{c} \frac{\partial \vec{D}}{\partial t}, \tag{2}$$

as well as the three "constitutive equations" (involving, respectively, $\vec{D}$ and $\vec{E}$, $\vec{J}$ and $\vec{E}$, and $\vec{B}$ and $\vec{H}$, and some of their derivatives) which characterize a given ferromagnetic medium. It may be noted that the divergence equations have not been listed because $\vec{\nabla} \cdot \vec{D} = 4\pi\rho$ follows from (2) if charge conservation is assumed, and $\vec{\nabla} \cdot \vec{B} = 0$ is an immediate consequence of (1). The constitutive equations contain certain material "constants," the permittivity, conductivity, and some generalized "permeability," and it will be seen later on that it is convenient to define these "constants" by the constitutive equations in which they occur. This procedure is always used in practice but it is logically unsatisfactory. The constitutive equations do not really define the "constants" since they already act as auxiliary relations for making the solutions of Maxwell's equations unique. However, it is well known that this difficulty can be resolved.[7] One can define the vectors $\vec{E}$ and $\vec{B}$ as certain averages of the microscopic electric and magnetic fields, respectively; the vectors $\vec{P}$ and $\vec{M}$ as the number of atoms per unit volume times certain averages of the atomic electric and magnetic moments, respectively; and the vector $\vec{J}$ as the number of conduction electrons per unit volume times a certain average of their velocities. Since $\vec{D}$ and $\vec{H}$ can be defined by the usual relations

$$\vec{D} = \vec{E} + 4\pi\vec{P}, \tag{3}$$
$$\vec{H} = \vec{B} - 4\pi\vec{M}, \tag{4}$$

It follows that $\vec{H}, \vec{B}, \vec{E}, \vec{D}, \vec{J}$, as well as the various material "constants," are now uniquely defined in principle.

At any given point the above-mentioned averages should be taken over a volume V whose linear dimensions

are large compared to interatomic distances but small compared to a distance L in which the ordinary field vectors change appreciably. When applied to ferromagnetic media, the latter qualification clearly requires that L does not exceed the smallest dimension, near the point in question, of any of the following possible inhomogeneities: a macroscopic inclusion, cavity, or stress variation, a crystallite, or a domain or a domain wall. The requirement that the linear dimensions of V be small compared to the thickness of a domain wall is particularly important. It shows that the "ordinary field vectors" do not average out the exchange effects arising from a rapid (but macroscopic) spatial variation of the direction of magnetization. In contrast the "average field vectors," which will be introduced later, do average out such exchange effects in domain walls but not in all other situations.

Except in unusual cases, which are mentioned below and discussed in Appendix I, all the known electromagnetic properties of ferromagnetic media can be represented by constitutive equations of the form

$$\vec{D} = \epsilon\vec{E}, \tag{5}$$
$$\vec{J} = \sigma\vec{E}, \tag{6}$$
$$\vec{G}(\vec{M}, \partial\vec{M}/\partial t, \nabla^2\vec{M}, \vec{H}) = 0, \tag{7}$$

where the complex permittivity ($\epsilon = \epsilon_1 - i\epsilon_2$), as well as the real conductivity σ, is a scalar which is independent of M and H. The vector function $\vec{G}$ will be specified later; for the sake of simplicity it is written in terms of $\vec{M}$ and $\vec{H}$, but it could be written in terms of $\vec{B}$ and $\vec{H}$ by making use of (4). The form indicated for $\vec{G}$ is sufficiently general to represent the cases where (due to the exchange effect) $\vec{G}$ is a function not only of $\vec{M}$ and $\vec{H}$ but of $\nabla^2\vec{M}$ as well. It should be stated that (6) neglects the Hall and magneto-resistance effects, and (5), (6), and (7) assume that the medium possesses a cubic crystal structure.

C. The Average Field Vectors and their Interrelations

(c) Only those physical situations will be considered in this paper which permit one to define the "average field vectors" introduced below.

Let $\vec{F}$ denote any of the vectors $\vec{H}, \vec{B}, \vec{E}, \vec{D}, \vec{J}, \vec{P}, \vec{M}$, and let $\vec{f}(\vec{f} = \vec{h}, \vec{b}, \cdots, \vec{m})$ denote its harmonically time dependent component. Assume that it is possible, at any point in the medium, to find a volume V' which encloses this point and possesses the following two properties: First, the linear dimension of V' in the direction of any specified unit vector $\vec{r}_0$ is large compared to the largest dimension, near the point in question, of any of the inhomogeneities (domains, etc.) mentioned in Section IIB. Second, this linear dimension of V' is small compared to $\lambda/4$; the wavelength λ of $\langle\vec{f}\rangle$, a vector to be defined later, is that wavelength which corresponds to the component parallel to $\vec{r}_0$ of the average propagation vector in the medium. The average

[7] J. H. Van Vleck, "The Theory of Electric and Magnetic Susceptibilities," Oxford University Press, London, pp. 1–13; 1932.

vectors $\langle\vec{F}\rangle$ and $\langle\vec{f}\rangle$ at the point in question are then defined to be the spatial averages over V' of $\vec{F}$ and $\vec{f}$, respectively.

It should be mentioned that this definition of the average vectors may appear to involve a "circular argument" because the specification of the upper limit of V' made use of the wavelength of $\langle\vec{f}\rangle$, and of the average propagation vector, at a time when $\langle\vec{f}\rangle$ had not yet been defined. However, this difficulty will not occur in practice because one can always assume some sufficiently large linear dimension of V', say 100 times the size of the known inhomogeneities, then take λ from experiment (or from a suitable averaging calculation involving the ordinary field vectors), and finally investigate whether the assumed linear dimension of V' is sufficiently small compared to $\lambda/4$.

Several consequences of the above definitions of the average vectors will now be noted. As implied in Section IIB, the vectors $\langle\vec{F}\rangle$ and $\langle\vec{f}\rangle$ do, by definition, average out the exchange effects in domain walls, but they do not average out the exchange effects when the latter are not localized, as in the situation considered in Section IV. Another consequence of the above definition is the fact that, because of the skin effect, a different shape of V' must be used in good conductors (e.g., metals), than in poor conductors (e.g., ferrites); in the former case V' is evidently shaped like a thin "pill box" parallel and adjacent to the surface of the medium, while in the latter case V' is roughly spherical throughout the medium. It should also be stated that the average vectors introduced above do not always exist. For example, in the case of a medium possessing "relatively" large crystallites, the kind of V' specified above may not exist everywhere, and hence $\langle\vec{F}\rangle$ and $\langle\vec{f}\rangle$ cannot be defined. On the other hand, it is sometimes possible to define $\langle\vec{F}\rangle$ and $\langle\vec{f}\rangle$ in certain special cases such as in uniformly stressed media which will nevertheless be excluded from the present considerations. The assumptions required for this exclusion will be stated in Section IIG and Section IIH; they concern the meaning of "anisotropy" and "homogeneity" of a medium with respect to the average field vectors.

It follows from the linearity of the Maxwell equations (1) and (2) that the average vectors also satisfy these equations. Thus the equations

$$\vec{\nabla} \times \langle\vec{E}\rangle = -\frac{1}{c}\frac{\partial\langle\vec{B}\rangle}{\partial t}, \tag{8}$$

$$\vec{\nabla} \times \langle\vec{H}\rangle = \frac{4\pi}{c}\langle\vec{J}\rangle + \frac{1}{c}\frac{\partial\langle\vec{D}\rangle}{\partial t}, \tag{9}$$

are always valid as long as the $\langle\vec{F}\rangle$ can be defined. However, the form of the "average" constitutive equations

$$\langle\vec{D}\rangle = \epsilon\langle\vec{E}\rangle, \tag{10}$$

$$\langle\vec{J}\rangle = \sigma\langle\vec{E}\rangle, \tag{11}$$

$$\vec{G'}(\langle\vec{M}\rangle,\ \partial\langle\vec{M}\rangle/\partial t,\ \nabla^2\langle\vec{M}\rangle,\langle\vec{H}\rangle) = 0, \tag{12}$$

requires further comments. Eq. (11) neglects the Hall and magnetoresistance effects, and (10), (11), and (12) assume that the medium possesses a cubic crystal structure. Only when these conditions are satisfied and the medium is homogeneous with respect to the ordinary $\vec{E}$, are the ϵ and σ in (10) and (11) the same as the ϵ and σ in (5) and (6), respectively. It should further be emphasized that the vector function $\vec{G'}$ in (12) is generally different from the $\vec{G}$ in (7); this statement will be illustrated by the examples of the next Section.

D. Necessity for Distinguishing between the Ordinary and Average Field Vectors

(d) It is assumed that whenever $\vec{F}$ and $\langle\vec{F}\rangle$ are not identical, $\vec{F}$ must be used in any (classical) calculation which derives the ordinary or the average constitutive equations from the basic physical properties of the medium; on the other hand, $\langle\vec{F}\rangle$ and particularly its alternating component $\langle\vec{f}\rangle$ is necessarily used in the evaluation of any electromagnetic experiment that involves time-dependent fields.

The validity of (d) may now seem evident because $\vec{F}$ and $\langle\vec{F}\rangle$ have been carefully distinguished in the preceding discussion. However, the writer believes that (d) is sufficiently important to justify amplification by means of the following four examples.

1. In a partially magnetized (i.e., nonsaturated) polycrystalline ferrite the relation between $\vec{b}$ and $\vec{h}$ under the assumptions of this paper, is given by a spin wave equation within the domain walls but by a tensor inside the domains; however, the (experimentally important) relation between $\langle\vec{b}\rangle$ and $\langle\vec{h}\rangle$ is given by a (perhaps position dependent) tensor everywhere (Section IIIA). As shown in Section IIIB2 the latter tensor can sometimes be calculated, but this is done by using $\vec{b}$, $\vec{h}$, etc., and averaging the relation existing between them.[3]

2. In the special case where the ferrite of example 1 has zero static magnetization, the tensor relating $\langle\vec{b}\rangle$ to $\langle\vec{h}\rangle$ reduces to a complex scalar permeability whose value is constant throughout the medium. When this (initial) permeability is measured as a function of frequency (yielding the so-called magnetic spectrum), one obtains one or two "natural resonances."[8] The enormously large width of these resonances is definitely not a measure of the basic damping processes involved in the $\vec{b}$ versus $\vec{h}$ relation, but rather a measure of the distribution[8] of resonance frequencies in a polycrystal. As emphasized by the writer[8] earlier, this distribution is "hidden" by the unavoidable representation of the data in terms of a $\langle\vec{b}\rangle$ vs $\langle\vec{h}\rangle$ relation.

3. At megacycle frequencies the initial permeability of a polycrystalline ferromagnetic metal, like that of the

[8] G. T. Rado, "Magnetic spectra of ferrites," Rev. Modern Phys., vol. 25, pp. 81–89; January, 1953.

rite of example 2, is described by a complex scalar. a metal, however, the imaginary part of this permea- ty is not solely a measure of relaxation losses but tially a measure of Joule heating, i.e., eddy current sipation.[9] These so-called microscopic eddy current ses, which are fundamentally not different from croscopic eddy current losses, are always regarded as ing part of the specifically magnetic losses because the periments are interpreted in terms of the average tors $\langle \vec{b} \rangle$, $\langle \vec{h} \rangle$, etc. These average vectors are much too barse" to take into account the domain structure of e metal, and thus they cannot be used for calculating e total eddy current losses. On the other hand, one uld (in principle) calculate a position dependent ation between $\vec{b}$ and $\vec{h}$ from the physical properties d the domain structure of the metal. Using the linary field vectors one could then calculate the total dy current loss by the third term of (62) of Appendix If this procedure could actually be carried out, then e specifically "magnetic" losses (as calculated from the st term of (62)) would not contain any eddy current tribution and would really be specifically magnetic ses. It is seen, therefore, that the magnitude of what ction of the total loss one chooses to call "eddy rrent loss" depends on whether ordinary or average d vectors are used in the calculation.

The situation described above is clearly a consequence the domain structure of the metal. If the metal is agnetically saturated so that the domain structure sappears, then there is no longer a distinction between e ordinary and average field vectors. In that case, erefore, the imaginary parts of the appropriate (tensor) rmeability do not even partially represent eddy current ects. This was first pointed out by Kittel[10] in connec- n with ferromagnetic resonance. In spin wave reso- nce (see Section IV), of course, one again encounters idden" eddy current losses but these occur not be- use the experiments are described by the average ctors (which does not cause any complications since e metal is saturated) but because of the presence of change effects.

4. At kilocycle frequencies the initial permeability of a lycrystalline ferromagnetic metal is generally a scalar lating $\langle \vec{b} \rangle$ to $\langle \vec{h} \rangle$) which is usually assumed to be real ther than complex. In contrast to example 3, there is erefore no ambiguity as to what is meant by eddy rrent losses. However, the calculated eddy current ses are invariably smaller than the measured losses. is discrepancy is known as the "eddy current anom- y." Apart from the fact that the measured losses metimes include relaxation losses, the eddy current omaly is probably due to the assumption of a real alar permeability. If the permeability were assumed

to be complex, then the correct total losses could un- doubtedly be calculated by (67). Although the latter equation involves the average field vectors so that the losses obtained from it are approximate rather than exact, as explained in Appendix II, in the present case it should yield exact results because the complex initial permea- bility is in the first place deduced from experiment; this can be done, for example, by measuring not only the inductance but also the resistance of a coil containing a ferromagnetic core. It is seen, therefore, that in this case the use of (67) merely amounts to a consistency check. A fundamental calculation of losses would have to be based on (62), but this is usually too difficult in practice; as in example 3, such a calculation requires, among other things, a detailed knowledge of the domain structure.

An interpretation of the eddy current anomaly was first given by Williams, Shockley, and Kittel;[11] their arguments are somewhat different from those given above but the two viewpoints are probably equivalent. In addition, these authors showed convincingly that in a simple domain structure the measured eddy current losses associated with domain wall motion can be calculated correctly by using the ordinary field vectors. Thus there is no anomaly when such a calculation can actually be carried out.

E. Frequency Range of the Time Dependent Field Vectors

(e) It is assumed that each vector $\vec{f}$ has the form $\vec{f} = \vec{f_o} \exp (i\omega t)$, where $\vec{f_o}$ is complex so that the phase of $\vec{f_o}$ may be arbitrary and the phases of its rectangular components may all be different; it is further assumed that the frequency $\omega/2\pi$ does not exceed 10^5 mega- cycles/second, corresponding to a free-space wavelength of 3 mm.

Although this frequency limit is very approximate, and may perhaps be increased by a factor of ten or more, it does not seem useful at present to discuss those "exchange" phenomena which pertain specifically to the infrared region. Such phenomena have been studied theoretically[12] but not yet experimentally.

Another reason for (e) is the following: if wavelengths much smaller than the above-mentioned limit were to be used, then it would not usually be possible (even in sintered ferrites) to define the average vectors intro- duced in Section IIC.

F. Linearity and Homogeneity of the $\vec{b}$ versus $\vec{h}$ Relation

(f) It is assumed that the equation relating $\vec{b}$ to $\vec{h}$ is linear and homogeneous in $\vec{b}$, $\vec{h}$, and their spatial derivatives.

[9] C. Kittel, "Ferromagnetic resonance," *Jour. Phys. Radium*, . 12, pp. 291–302; March, 1951.
[10] C. Kittel, "On the theory of ferromagnetic resonance absorp- n," *Phys. Rev.*, vol. 73, pp. 155–161; January, 1948.

[11] H. J. Williams, W. Shockley, and C. Kittel, "Studies of the propagation velocity of a ferromagnetic domain boundary," *Phys. Rev.*, vol. 80, pp. 1090–1094; December, 1950.
[12] J. Kaplan and C. Kittel, "Exchange frequency electron spin resonance in ferrites," *Jour. Chem. Phys.*, vol. 21, pp. 760–761; April, 1953.

The validity of the linearity assumption in the usual practical situations has been verified by numerous experiments, and can be understood theoretically by considering particular physical models (see Section IIIB) of a ferromagnetic medium. However, it should be mentioned that at least one series of experiments[13] has demonstrated, in confirmation of magnetic theory, that for sufficiently large values of $\vec{h}$ the relation between $\vec{b}$ and $\vec{h}$ is nonlinear. Despite its potential interest, such a nonlinear situation occurs relatively rarely at microwave frequencies because ordinarily a large amount of power is required to produce it experimentally. Furthermore, Maxwell's equations do not seem to have been solved as yet for those cases where $\vec{b}$ and $\vec{h}$ are not parallel and are not linearly related to each other.

The validity of the homogeneity assumption follows from physical considerations. Considering only the nontrivial case where $\vec{b} \neq \vec{h}$, one can argue that if the homogeneity assumption were not fulfilled, then $\vec{b}_o$ would not necessarily vanish if $\vec{h}_o$ vanishes. But this would mean that no energy dissipation takes place, in contradiction to the known fact that in all ferromagnetic media some losses occur. Thus the homogeneity assumption must generally be valid.

It should also be mentioned that, as a consequence of (f), the equation relating $\langle \vec{b} \rangle$ to $\langle \vec{h} \rangle$ is necessarily linear and homogeneous in $\langle \vec{b} \rangle$, $\langle \vec{h} \rangle$, and their spatial derivatives.

G. *Meaning of Anisotropy with Respect to the Average Field Vectors*

(g) It is assumed that the medium possesses at every point: (1) isotropy with respect to the static component of $\langle \vec{F} \rangle$; and (2) either isotropy or uniaxial anisotropy with respect to $\langle \vec{f} \rangle$, the anisotropy axis being determined by the direction of the static component of $\langle \vec{M} \rangle$ or $\langle \vec{H} \rangle$ at the point in question.

Part (1) of (g) means, in particular, that the average static magnetic properties in the vicinity of a point do not depend on direction. Thus the static $\langle \vec{M} \rangle$ and the static $\langle \vec{H} \rangle$ are parallel if they both exist. However, their magnitudes and directions may be position-dependent; this somewhat surprising possibility is clarified in Section IIH.

This isotropy assumption is introduced to exclude from the present considerations such special cases as single crystals and uniformly stressed media, neither of which are necessarily excluded by assumption (c). Although the contents of Sections III and IV could undoubtedly be extended by means of standard domain theory to apply to these special cases, such an extension is not yet necessary because the corresponding wave

[13] N. Bloembergen and S. Wang, "Relaxation effects in para— and ferromagnetic resonance," *Phys. Rev.*, vol. 93, pp. 72–83; January, 1954.

propagation experiments have been performed only single crystals which are small compared to $\lambda/4$; in su a situation the wave propagation problem reduces to quasi-static problem[10] and the tensor description of t permeability is not essential.

Part (2) of (g) is a direct consequence of part (1) a of assumption (c). For if the average vectors can defined, then at any point in the medium the on possible preferred direction (from the "point of vie of $\langle \vec{f} \rangle$) is that of the static component of $\langle \vec{M} \rangle$ or $\langle \vec{H} \rangle$ The anisotropy axis is, in general, determined by t common direction of these vectors; at the remanence a coercivity points, however, the anisotropy axis is det mined by the direction of the static component of $\langle \vec{M} \rangle$ and $\langle \vec{H} \rangle$, respectively.

H. *Meaning of Homogeneity with Respect to the Avera Field Vectors*

(h) It is assumed that the medium is homogened with respect to the static component of $\langle \vec{F} \rangle$ but n necessarily homogeneous with respect to $\langle \vec{f} \rangle$.

This assumption is not independent and is stated he merely for clarity. Since (c) asserts that the mediu under consideration permits the average vectors to defined, it follows that the medium is necessari homogeneous with respect to the static component $\langle \vec{F} \rangle$. This means that the "average" static properties the medium do not vary from point to point.

However, the medium is not necessarily homogeneo with respect to $\langle \vec{f} \rangle$. This is due to the fact that the tim dependent properties of the medium depend on i static "state" rather than on its static properties. F example, the static $\langle \vec{H} \rangle$ (and hence the static $\langle \vec{M} \rangle$ ar $\langle \vec{B} \rangle$) may vary from point to point in both directi and magnitude; this could be due to the applied fie being nonuniform or the sample shape being irregula In such cases the equation relating $\langle \vec{b} \rangle$ to $\langle \vec{h} \rangle$ mu evidently vary from point to point in the mediu even though the vectors $\langle \vec{B} \rangle$ and $\langle \vec{H} \rangle$ [which a parallel due to (g)] have the same ratio everywhe While in some practical situations the medium is in fa homogeneous with respect to $\langle \vec{f} \rangle$ also, a circumstan which greatly simplifies the electromagnetic calculatio it should be emphasized that even in saturated med this is not always the case.

III. Equations that Connect $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ by a Point Relation

Throughout Section III it will be assumed tha
(i) At every point in the medium the vectors $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ are connected by a point relation. By definition, such relation does not contain any spatial derivatives.

A. *General Form of the Permeability Tensor*

The most general relation between $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ whic is consistent with the assumptions (a) through (i), b

s not yet take advantage of the second part of (g), necessarily given by the equations

$$\langle b_x \rangle = \mu_{11} \langle h_x \rangle + \mu_{12} \langle h_y \rangle + \mu_{13} \langle h_z \rangle, \qquad (13a)$$
$$\langle b_y \rangle = \mu_{21} \langle h_x \rangle + \mu_{22} \langle h_y \rangle + \mu_{23} \langle h_z \rangle, \qquad (13b)$$
$$\langle b_z \rangle = \mu_{31} \langle h_x \rangle + \mu_{32} \langle h_y \rangle + \mu_{33} \langle h_z \rangle, \qquad (13c)$$

ere the subscripts, x, y, z, designate the components b) and $\langle h \rangle$ along the axes of an arbitrary rectangular ordinate system, and the quantities μ_{jk} (j, k = , 3) are coefficients to be discussed below. Since the represent "average" permeabilities in the sense of tion IIC, they should have been written as $\langle \mu_{jk} \rangle$, for simplicity the angular parentheses are omitted. Evidently the vector $\langle \vec{b} \rangle$, which is ultimately obtained er subjecting the given vector $\langle \vec{h} \rangle$ to the "com- ent operations" involved in (13), must be independent the orientation of the co-ordinate system used. nsequently the matrix $\| \mu \|$ of the coefficients μ_{jk} ot just some matrix that relates an arbitrary set of mbers $\langle h_x \rangle$, $\langle h_y \rangle$, $\langle h_z \rangle$ to another arbitrary set of mbers $\langle b_x \rangle$, $\langle b_y \rangle$, $\langle b_z \rangle$. Instead, $\| \mu \|$ represents the trix of the components (in some arbitrary coordinate tem) of a certain physical quantity called the permea- ity tensor $\bar{\bar{\mu}}$. This tensor is defined to be the operator t assigns at every point a certain vector $\langle \vec{b} \rangle$ to the tor $\langle \vec{h} \rangle$ in a homogeneous and linear fashion. To ress the role of this operator symbolically, one tes the equation

$$\langle \vec{b} \rangle = \bar{\bar{\mu}} \langle \vec{h} \rangle. \qquad (14)$$

make actual calculations, however, one uses not the antities appearing in (14) but the matrices which respond to them in some arbitrary co-ordinate system. r example, (13) can be obtained as the result of the trix multiplication corresponding to the operation earing in (14).

Since $\| \mu \|$ is a matrix of tensor components, it must ssess certain definite transformation properties with pect to a rotation of the co-ordinate system. These perties lead to certain invariances (namely the trix' symmetry or antisymmetry, its trace, and its ular equation), but they do not lead to invariance of e whole matrix. Thus the tensor $\bar{\bar{\mu}}$ is, by definition, ariant with respect to a rotation of the co-ordinate tem, but the matrix $\| \mu \|$ is not invariant with pect to such a rotation.

Before discussing $\bar{\bar{\mu}}$, it should be noted that the second rt of assumption (g) makes it possible to simplify th $\bar{\bar{\mu}}$ and a certain representation of $\| \mu \|$. To do this, ose any point in the medium and let $\vec{i_z}$ be a unit ctor parallel to the static component of the $\langle \vec{H} \rangle$ or) that defines the axis of the anisotropy introduced (g). Now consider two vectors $\langle \vec{h_1} \rangle$ and $\langle \vec{h_2} \rangle$ which are ual in magnitude and oriented arbitrarily with respect $\vec{i_z}$; evidently $\bar{\bar{\mu}}$ assigns to these two vectors certain ctors $\langle \vec{b_1} \rangle$ and $\langle \vec{b_2} \rangle$, respectively, which do not neces-

sarily have any simple relation to $\langle \vec{h_1} \rangle$ and $\langle \vec{h_2} \rangle$ or to $\vec{i_z}$. But if $\langle \vec{h_1} \rangle$ and $\langle \vec{h_2} \rangle$ are not only equal in magnitude but form equal angles with $\vec{i_z}$, then it follows from (g) that $\bar{\bar{\mu}}$ must have the following symmetry property: the $\langle \vec{b_1} \rangle$ and $\langle \vec{b_2} \rangle$ resulting from the tensor operation must be equal in magnitude, and $\langle \vec{b_1} \rangle$ must have the same spatial relation to $\langle \vec{h_1} \rangle$ and $\vec{i_z}$ as $\langle \vec{b_2} \rangle$ has to $\langle \vec{h_2} \rangle$ and $\vec{i_z}$. Since all the five vectors under discussion start from a common point, this symmetry property of $\bar{\bar{\mu}}$ simply means that the pyramids formed by $\langle \vec{h_1} \rangle$, $\langle \vec{b_1} \rangle$, $\vec{i_z}$, and $\langle \vec{h_2} \rangle$, $\langle \vec{b_2} \rangle$, $\vec{i_z}$, are congruent.

Now choose some rectangular coordinate system whose positive z-direction is parallel to $\vec{i_z}$. The above mentioned symmetry property clearly requires that the whole matrix $\| \mu \|$ of the tensor components must remain unchanged if the coordinate system is rotated around the z-axis. An elementary calculation shows that this invariance is possible only if the tensor components μ_{jk} satisfy the relations

$$\mu_{11} = \mu_{22}, \qquad (15a)$$
$$\mu_{12} = -\mu_{21}, \qquad (15b)$$
$$\mu_{13} = \mu_{23} = \mu_{31} = \mu_{32} = 0. \qquad (15c)$$

Since it is convenient to introduce the notation

$$\mu_{11} = \mu, \qquad (16a)$$
$$\mu_{12} = -i\mu_t, \qquad (16b)$$
$$\mu_{33} = \mu_z, \qquad (16c)$$

it is seen from (15) that for the type of co-ordinate system under discussion the matrix $\| \mu \|$ has the form

$$\| \mu \| = \begin{Vmatrix} \mu & -i\mu_t & 0 \\ i\mu_t & \mu & 0 \\ 0 & 0 & \mu_z \end{Vmatrix}. \qquad (17)$$

This general result will now be discussed.

1. It follows from the assumptions introduced in this paper that the form of the matrix (17) is valid even if the medium is magnetically nonsaturated; thus (17) is more general than Polder's[4] matrix which is, even in its general form, valid only if the medium is saturated.

2. As discussed in Sections IIG and IIH, the direction of $\vec{i_z}$ of the axis of $\bar{\bar{\mu}}$, as well as the magnitudes of the tensor components, may vary from point to point in the medium; these properties are possessed by Polder's tensor also, but in his case the variations are due solely to variations in the static part of $\vec{H}$.

3. In general, the tensor components must be complex numbers in order to allow a representation of magnetic losses. Thus μ, μ_t, and μ_z must have the form

$$\mu = \mu_1 - i\mu_2, \qquad (18a)$$
$$\mu_t = \mu_{t1} - i\mu_{t2}, \qquad (18b)$$
$$\mu_z = \mu_{z1} - i\mu_{z2}. \qquad (18c)$$

The physical requirement that the losses be non-nega- tive, *i.e.*, that the time-average energy absorption by the medium be positive or zero, imposes the conditions

$$\mu_2 \geq 0, \tag{19a}$$
$$|\mu_{t2}| \leq \mu_2, \tag{19b}$$
$$\mu_{z2} \geq 0, \tag{19c}$$

on the tensor components; if the medium is lossless, then the relations (19) are equalities. It is seen that the requirement of non-negative losses determines the sign of μ_2 and μ_{z2}, but not the sign of μ_{t2}. To prove (19), one can use the expression

$$P_m = (1/4\pi)\, \vec{\langle h \rangle} \cdot \partial \vec{\langle b \rangle}/\partial t \tag{20}$$

for the "magnetic" contribution to the energy absorbed per second per cm³ by the medium from the field. Eq. (20) applies in the vicinity of an arbitrary point in the medium and would have to be integrated over the volume of a given specimen to obtain the total magnetic loss. As shown in Appendix II, (20) is not rigorous but usually it represents a fair approximation. However, (20) does hold rigorously if the relation between $\vec{\langle b \rangle}$ and $\vec{\langle h \rangle}$ (which is a tensor in the situations considered in the present section) is obtained from experiment. This was discussed in example 4 of Section IID. Since $\vec{\langle b \rangle}$ and $\vec{\langle h \rangle}$ are complex vectors and have a harmonic time dependence, the time average of P_m, denoted by $\langle\langle P_m \rangle\rangle$, may be written in the form

$$\langle\langle P_m \rangle\rangle = (1/8\pi)\, \mathrm{Re}\,\{\vec{\langle h \rangle}^* \cdot \partial \vec{\langle b \rangle}/\partial t\}$$
$$= -(\omega/8\pi)\, \mathrm{Im}\,\{\vec{\langle h \rangle}^* \cdot \vec{\langle b \rangle}\}, \tag{21a}$$

which may be combined with the general equation

$$\vec{\langle b \rangle} = \vec{\langle h \rangle} + 4\pi \vec{\langle m \rangle} \tag{22}$$

to yield the equivalent form

$$\langle\langle P_m \rangle\rangle = -(\omega/2)\, \mathrm{Im}\,\{\vec{\langle h \rangle}^* \cdot \vec{\langle m \rangle}\}. \tag{21b}$$

The asterisk (*) is used throughout this paper to denote the complex conjugate of the quantity to which it is applied; in (21a) and (21b) the asterisk could evidently have been placed on $\vec{\langle b \rangle}$ and $\vec{\langle m \rangle}$, respectively, instead of on $\vec{\langle h \rangle}$. When (21a) is used to prove (19b), it is convenient to introduce the vector $\vec{\langle h \rangle}_t$ which is defined to be the vector component of $\vec{\langle h \rangle}$ in the plane perpendicular to $\vec{i}_z$. But in the course of this proof[14] it must be remembered that $\vec{\langle h \rangle}_t$ is complex so that the vector product

$$\vec{\langle h \rangle}_t \times \vec{\langle h \rangle}_t^*$$

does not generally vanish. It is useful to express $\vec{\langle h \rangle}_t$ in the form

$$\vec{\langle h \rangle}_t = e^{ia}(\vec{\langle h_1 \rangle} - i\vec{\langle h_2 \rangle}),$$

where $\vec{\langle h_1 \rangle}$ and $\vec{\langle h_2 \rangle}$ are real, orthogonal, vectors in the plane perpendicular to $\vec{i}_z$. If one then combines (21a) with (24b), the relation (19b) is easily obtained.

4. In the case of local uniaxial symmetry [see assumption (g)] around $\vec{i}_z$, the role of the tensor $\overline{\overline{\mu}}$ can be ex-

pressed by using the $\|\mu\|$ of (17), so that one obtains

$$\langle b_x \rangle = \mu \langle h_x \rangle - i\mu_t \langle h_y \rangle, \tag{23}$$
$$\langle b_y \rangle = i\mu_t \langle h_x \rangle + \mu \langle h_y \rangle, \tag{23}$$
$$\langle b_z \rangle = \mu_z \langle h_z \rangle, \tag{23}$$

where the mutually orthogonal x-, y-axes are, of course arbitrarily oriented within the plane perpendicular $\vec{i}_z$. Alternatively, the role of $\overline{\overline{\mu}}$ can be expressed by t equation

$$\vec{\langle b \rangle} = \mu \vec{\langle h \rangle} + i\mu_t \vec{i}_z \times \vec{\langle h \rangle} + (\mu_z - \mu)(\vec{\langle h \rangle} \cdot \vec{i}_z)\vec{i}_z. \tag{24a}$$

Using the $\vec{\langle h \rangle}_t$ defined above, (24a) can be written in t form

$$\vec{\langle b \rangle} = \mu \vec{\langle h \rangle}_t + i\mu_t \vec{i}_z \times \vec{\langle h \rangle}_t + \mu_z \langle h_z \rangle \vec{i}_z. \tag{24b}$$

5. In the limiting case where the medium is ma netically saturated, the reversible "parallel perme bility" along $\vec{i}_z$ must be unity irrespective of any physi model. Thus one obtains

$$\mu_z = 1$$

for this case, so that (17) properly reduces to Polder general result. In the other limiting case, that of medium which is demagnetized, the static field a static magnetization are both zero. Then one obtains the basis of symmetry considerations alone

$$\mu_t = 0$$
$$\mu_z = 0$$

for this case, so that $\overline{\overline{\mu}}$ reduces to the isotropic init permeability $\mu = \mu_1 - i\mu_2$.

6. From (17) and (22) it follows that the mat $\|x\|$ of the susceptibility tensor components must ha the form of (17), and that the components of $\overline{\overline{x}}$ a related to the corresponding components of $\overline{\overline{\mu}}$ by t relations

$$x = (\mu - 1)/4\pi, \tag{25}$$
$$\pm i x_t = \pm i\mu_t/4\pi, \tag{25}$$
$$x_z = (\mu_z - 1)/4\pi. \tag{25}$$

B. Special Forms of the Permeability Tensor

In this section the tensor components μ, $\pm i\mu_t$, a μ_z will be calculated on the basis of special physic models. The three cases to be considered are: (1) s urated ferromagnetic media; (2) arbitrarily magnetiz ferromagnetic media; and (3) arbitrarily magnetiz ferrimagnetic media. These three cases, first treated Polder,[4] the writer,[3] and Wangsness,[5] respectively, w be reviewed from the standpoint of the present pap Thus the terminology and notation differ somewh from that used in the original papers.

1. *Saturated ferromagnetic media.* The assumptio made in Polder's theory of the tensor components (distinguished from his theory of the form of the tens will first be discussed. They will be compared, whenev possible, with the assumptions (a) through (i) which a too elaborate for the present purpose since they we ormulated, in Sections II and III*A*, mainly for t

[14] This method of proving the relation (19b) was suggested by M. L. Kales.

urpose of deriving the general form of the permeability
nsor.

In Polder's theory, assumption (a) is not stated
xplicitly, but the resulting tensor components do apply[5]
ferrimagnetic media if there are no "compensation
bints." An assumption equivalent to (b) is used, but
) is replaced by the equation of motion (26) discussed
elow. Assumptions (c) and (d) do not occur because
older treats only saturated media so that there is no
istinction between the ordinary and average vectors.
ssumption (e) is made, although the upper limit for
he frequency is not specified explicitly. Assumption (f)
automatically fulfilled by the equation of motion (26)
nd the "linearizing" assumptions (28) discussed below.
nstead of (g), the assumption is made that the medium
magnetically saturated. Thus there is uniaxial aniso-
ropy with respect to $\vec{f}$ at every point, the axis of this
nisotropy (i.e., the z-axis) being in the direction of the
tatic magnetic field $\vec{H}_o$ at the point in question; both
he magnitude and direction of $\vec{H}_o$ are allowed to vary
rom point to point in the medium. Magnetocrystalline
nisotropy and stress anisotropy are neglected; this
heans, in effect, that $\vec{H}_o$ is assumed to be sufficiently
arge, and that single crystals are excluded. Instead of
h), it is assumed, in effect, that the medium is homo-
eneous with respect to the static component of $\vec{F}$ but
ot necessarily homogeneous with respect to $\vec{f}$. The
atter part of this statement allows for the possibility
see Section IIH) that the direction $(\vec{i}_z)$ and magnitude
f $\vec{H}_o$, and thus the direction $(\vec{i}_z)$ of $\vec{M}_s$ (the static part
f $\vec{M}$) may vary from point to point even though the
nedium is saturated. Instead of (i), the more specific
ssumption is made that exchange effects due to spatial
ariations in the direction of $\vec{M}$ are negligible, and that
$\vec{M}$ and $\vec{H}$ are related by the classical equation of motion

$$(1/\gamma)\partial\vec{M}/\partial t = \vec{M} \times \vec{H}, \qquad (26)$$

vhere γ is the magnetomechanical ratio. In terms of
tandard notation, γ is given by $ge/2mc$; however, in
errimagnetic media the spectroscopic splitting factor g
hust be replaced by the effective value g_{eff}, as shown
y Wangsness.[5] Eq. (26) equates the time rate of change
f angular momentum density to the torque density.
inally, it is assumed that in the relations

$$\vec{H} = \vec{H}_o + \vec{h} = H_o\vec{i}_z + \vec{h}, \qquad (27a)$$
$$\vec{M} = \vec{M}_s + \vec{m} = M_s\vec{i}_z + \vec{m}, \qquad (27b)$$

he magnitudes of the periodic vectors $\vec{h}$ and $\vec{m}$ are
ufficiently small so that the approximations

$$h/H_o \ll 1, \qquad (28a)$$
$$m/M_s \ll 1, \qquad (28b)$$

re valid.

Substitution of (27) and (28) into (26), together with
he general equation

$$\vec{b} = \vec{h} + 4\pi\vec{m}, \qquad (29)$$

is easily seen to yield a tensor relation between $\vec{b}$ and
$\vec{h}$. In fact, $\| \mu \|$ is of the form (17), and μ, μ_t, μ_z are
given by

$$\mu = \frac{\gamma^2 H_o B_o - \omega^2}{\gamma^2 H_o^2 - \omega^2}, \qquad (30a)$$

$$\mu_t = \frac{4\pi M_s \gamma\omega}{\gamma^2 H_o^2 - \omega^2}, \qquad (30b)$$

$$\mu_z = 1, \qquad (30c)$$

where B_o denotes $H_o + 4\pi M_s$. Eqs. (30), which con-
stitute Polder's results, are seen to contain real numbers
only. This is due to the fact that (26) does not contain
any damping (or relaxation) term, and thus does not
account for energy dissipation. Such a term would limit
the amplitude of the forced motion of $\vec{M}$ and would, in
the absence of external excitation, cause $\vec{M}$ to return
to its equilibrium orientation $\vec{M}_s$. Thus the theory is not
really complete, especially since the existence of a
damping term must be assumed implicitly in order to
justify the approximations (28).

Two types of phenomenological damping term appear
in the literature. One type, due to Landau and Lifshitz,
has the form

$$-\frac{\lambda_r}{\gamma} \left[\frac{(\vec{H} \cdot \vec{M})\vec{M}}{M_s^2} - \vec{H} \right], \qquad (31)$$

where λ_r is a relaxation frequency. If the expression (31)
is added to the right hand side of (26) then μ_z is un-
changed by the damping, but μ and μ_t take on the
complex form (18a) and (18b), respectively. According
to Hogan,[15] the tensor components are then determined
by

$$\mu_1 = 1 + \frac{(\gamma^2 H_{oo}^2 - \omega^2)4\pi M_s\gamma^2 H_{oo}\chi_o^2 + 8\pi\omega^2\lambda_r^2\chi_o}{\chi_o^2(\gamma^2 H_{oo}^2 - \omega^2)^2 + 4\omega^2\lambda_r^2}, \qquad (32a)$$

$$\mu_2 = \frac{4\pi\lambda_r\omega\chi_o^2(\gamma^2 H_{oo}^2 + \omega^2)}{\chi_o^2(\gamma^2 H_{oo}^2 - \omega^2)^2 + 4\omega^2\lambda_r^2}, \qquad (32b)$$

$$\mu_{t1} = \frac{4\pi M_s\gamma\omega\chi_o^2(\gamma^2 H_{oo}^2 - \omega^2)}{\chi_o^2(\gamma^2 H_{oo}^2 - \omega^2)^2 + 4\omega^2\lambda_r^2}, \qquad (32c)$$

$$\mu_{t2} = \frac{8\pi\omega^2\gamma\lambda_r H_o\chi_o^2}{\chi_o^2(\gamma^2 H_{oo}^2 - \omega^2)^2 + 4\omega^2\lambda_r^2}, \qquad (32d)$$

where $\chi_o = M_s/H_o$ is the static susceptibility at the
point in question, and $H_{oo} = H_o[1 + (\lambda_r/\gamma M_s)^2]^{1/2}$ is
an abbreviation. It should be emphasized that neither
(31), nor an alternative damping term due to Bloch,
has been unambiguously verified experimentally. In
particular, it is not known to what extent λ_r depends on
H_o, ω, etc. Furthermore, in most media the required
magnitudes of the phenomenological relaxation fre-
quencies have not been explained theoretically.

---

[15] C. L. Hogan, "The ferromagnetic Faraday effect at micro-
wave frequencies and its applications," *Rev. Modern Phys.*, vol. 25,
pp. 253–263; January, 1953.

It will now be attempted to make plausible just why $\vec{H}$ is the field that occurs in (26). Clearly the field acting on a given atomic magnetic moment $\vec{p}_{at}$ cannot be $\vec{B}$, for it was stated in Section IIB that $\vec{B}$ is the average of the microscopic field, and thus $\vec{B}$ contains a dipolar contribution from $\vec{p}_{at}$ itself. Neither can the field acting on $\vec{p}_{at}$ be $\vec{H}$, for it is easily seen from Maxwell's (macroscopic) equations that, at least in the stationary state, $\vec{H}$ is the sum of two parts: one is due to currents, and the other arises from (surface and volume) divergences of $\vec{M}$. This means, for example, that if one imagines a uniform, unbounded, medium saturated in a uniform external field due to currents, then $\vec{H}$ is simply equal to this external field throughout the medium. In this special case, therefore, $\vec{H}$ does not contain a dipolar contribution from any atom whatsoever. (The essential aspects of this special case can be realized by the well known configuration of a uniform current winding covering a toroidal core.) Thus it is seen that even in the general case there are too many dipolar contributions contained in $\vec{B}$ and too few in $\vec{H}$. The true field acting on $\vec{p}_{at}$ (in the cubic lattices and locally aligned moments considered in the classical model of ferromagnetism) is given by $\vec{H} + (4\pi/3)\vec{M}$, which is smaller than $\vec{B}$ but larger than $\vec{H}$. The inclusion of the Lorentz local field $(4\pi/3)\vec{M}$ takes account of the fact that the true field must contain a contribution from the microscopic (dipole-dipole) fields due to all the atoms except the one atom at which $\vec{p}_{at}$ is located. If it is now remembered that the Weiss field $\vec{H}_w$ (which arises from the exchange effect) aligns the atomic moments approximately parallel to each other, so that $\vec{M}$ is given by $\vec{p}_{at}$ times the number of atoms per unit volume, and that $\vec{H}_w$ is parallel to $\vec{M}$, then it is seen that the equation

$$\vec{M} \times (\vec{H} + \vec{H}_w + 4\pi\vec{M}/3) = \vec{M} \times \vec{H}$$

is valid independent of the orientation of $\vec{M}$. Thus neither the Weiss field nor the Lorentz field occurs in (26). This conclusion was first arrived at by Kittel.[16] It may also be noted that while the physical content of Polder's theory is the same as that of Kittel's[16,10] theory, the Polder tensor formulation represents not only an alternative method for obtaining Kittel's results, but also a method for expressing the (magnetic) constitutive relation in a manner which is particularly suitable for use in connection with Maxwell's equations.

2. *Arbitrarily magnetized ferromagnetic media.* The calculation of the permeability tensor components has not yet been performed in sufficient generality to apply to all situations of experimental interest involving non-saturated media. However, the writer[3] has made an

approximate calculation which does apply to non-saturated media under some commonly encountered experimental conditions. In that calculation, assumption (a) is not stated explicitly, but the resulting tensor components do apply[5] to ferrimagnetic media if there are no "compensation points." Assumption (b) is used, but (7) is replaced by the equation of motion (33) discussed below. Assumptions (c) and (d) do necessarily occur; their formulations in the present paper represent, in fact, elaborations of the corresponding assumptions introduced in the writer's earlier calculation. Assumption (e) is made, although in the earlier calculation the upper limit for the frequency was implied rather than explicitly stated. Assumption (f) is automatically fulfilled by the equation of motion (33) and the "linearizing" assumptions (36) and (28b) discussed below. Assumptions (g) and (h) are both made. Instead of (i), the following more specific assumptions are introduced: (i_1) exchange effects due to spatial variations in the direction of $\vec{M}$ are negligible within domains; (i_2) the total volume of the domain walls is so small compared to the total volume of the domains that the contribution of the domain walls to the static bulk magnetization is negligible; (i_3) due to the use of a sufficiently high frequency, or for some other reason, the contribution of domain wall displacements to the time-dependent magnetization is negligible; (i_4) at points not contained within domain walls, $\vec{M}$ and $\vec{H}$ are related by the classical equation of motion

$$(1/\gamma)\partial\vec{M}/\partial t = \vec{M} \times \vec{H}', \qquad (33)$$

where γ (which contains g or g_{eff}) is the same as in (26). Before discussing $\vec{H}'$, which contains $\vec{H}$, it is convenient to introduce the time-independent unit vector $\vec{u}$ that describes the equilibrium orientation of $\vec{M}$ at the point in question. Since the magnetization inside a domain is saturated, $\vec{u}$ is defined by the equation

$$\vec{M} = \vec{M}_s + \vec{m} = M_s\vec{u} + \vec{m}, \qquad (34)$$

which takes the place of (27b). It should be emphasized that in a nonsaturated medium $\vec{u}$ is not the same as $\vec{i}_s$ because the latter points along the average direction of $\vec{M}_s$ over the averaging volume V'. Now introduce, at the point in question, an effective static field $\vec{H}_e$ which is responsible for $\vec{M}_s$ having the direction $\vec{u}$. The field

$$\vec{H}_e = H_e\vec{u}$$

may be thought of as being the resultant

$$\vec{H}_e = \vec{H}_o + \vec{H}_{ani} + \vec{H}_{str}$$

of the true static field $\vec{H}_o$ (which contains all static demagnetizing fields arising from the divergences of $\vec{M}_s$), the effective field due to anisotropy, $\vec{H}_{ani}$, and the effective field due to nonuniform stresses, $\vec{H}_{str}$. Although $\vec{H}_{ani}$ and $\vec{H}_{str}$ cannot be uniquely defined, it should be stated that for any given domain geometry

[16] C. Kittel, "Interpretation of anomalous Larmor frequencies in ferromagnetic resonance experiment," *Phys. Rev.*, vol. 71, pp. 270–271; February, 1947.

e torque densities $\vec{M} \times \vec{H}_{ani}$ and $\vec{M} \times \vec{H}_{str}$ can ways be uniquely defined (at least in principle) on e basis of the well known expressions for the free ergy due to anisotropy and stress, respectively. Thus ' is given by

$$\vec{H}' = \vec{H}_e + \vec{h} = H_e\vec{u} + \vec{h}, \qquad (35)$$

hich takes the place of (27a). It is evident from the oove discussion that, even within the averaging olume V' (see Section IIC), the quantities $\vec{m}$, $\vec{h}$, $\vec{u}$ and e are functions of position.

Although the effects of anisotropy, stress, and magtic interactions are included in $\vec{H}_e$, the calculation $\vec{H}_e$ (or rather of $\vec{M} \times \vec{H}_e$) is not usually possible a polycrystal. The problem encountered in such a lculation is not only mathematical but includes the fficulty of quantitatively specifying the nature of the ystallite structure, imperfections, and stresses. Thus a polycrystal the calculation of $\vec{H}_e$ involves the same rt of problems as the calculation of the domain ructure or the technical magnetization curve. Never-eless, one can establish a rough lower limit for H_e. a polycrystal (although not necessarily in a single ystal) H_e may be expected to exceed the larger one the two quantities H_o and K/M_s. Here K is the nisotropy constant, a quantity whose order of mag-tude is usually known. Consequently the approxima-on

$$h/H_e \ll 1, \qquad (36)$$

hich replaces (28a), is capable of being tested and is erefore meaningful even though H_e is not actually nown. Expressing $\vec{M}$ and $\vec{H}'$ by means of (34) and (35), spectively, and using the linearizing assumptions (36) nd (28b), one obtains from (33)

$$\vec{m} = (\gamma/i\omega)(H_e\vec{m} - M_s\vec{h})\times\vec{u}, \qquad (37)$$

hich is linear in $\vec{m}$ and $\vec{h}$ but still contains the unknown $_o$. To eliminate H_o, it is now necessary to make the urther approximations

$$H_e\gamma \ll \omega, \qquad (38)$$

$$4\pi M_s\gamma \ll \omega, \qquad (39)$$

hich are satisfied under many (but not all) experi-ental conditions. In order to be able to test (38), one eeds a rough upper limit for H_e. Such a limit is easily btained by noting that in a polycrystal H_e is expected be smaller than, or of the order of, that value of the rue static field which must be exceeded when saturating he medium. Inserting (38) and (39) into (37), one btains after a short calculation

$$\vec{b} \equiv \vec{h} + 4\pi\vec{m} \simeq \vec{h} + i(4\pi M_s\gamma/\omega)\vec{h} \times \vec{u}, \qquad (40)$$

here H_o no longer occurs. Eq. (40) shows clearly that he ordinary vectors $\vec{b}$ and $\vec{h}$ are related by a tensor. Iowever, the vector $\vec{u}$ is not known, so that (40) must e averaged over V' to obtain the (tensor) relation

between $\langle\vec{b}\rangle$ and $\langle\vec{h}\rangle$. To do this, write $\vec{u}$ and $\vec{h}$ in the form

$$\vec{u} = \langle\vec{u}\rangle + \delta\vec{u},$$
$$\vec{h} = \langle\vec{h}\rangle + \delta\vec{h},$$

respectively, so that $\langle\vec{u}\times\vec{h}\rangle$ becomes

$$\langle\vec{u}\times\vec{h}\rangle = \langle\langle\vec{u}\rangle\times\langle\vec{h}\rangle + \delta\vec{u}\times\delta\vec{h}\rangle = \langle\vec{u}\rangle\times\langle\vec{h}\rangle + \langle\delta\vec{u}\times\delta\vec{h}\rangle. \qquad (41)$$

The first equality in (41) arises from the fact that, by the definition of V', the averages $\langle\langle\vec{u}\rangle\times\delta\vec{h}\rangle$ and $\langle\langle\vec{h}\rangle\times\delta\vec{u}\rangle$ both must vanish. If it is now assumed that $\langle\delta\vec{u}\times\delta\vec{h}\rangle$ also vanishes, then (41) becomes

$$\langle\vec{u}\times\vec{h}\rangle = \langle\vec{u}\rangle\times\langle\vec{h}\rangle. \qquad (42)$$

To see what this means, choose some arbitrary point A within V' and let $\delta\vec{u}$ and $\delta\vec{h}$ form the angle α with each other. Since V' is much larger than a domain or cry-stallite, the above assumption asserts that there is some point B within V' at which $\delta\vec{u}$ and $\delta\vec{h}$ form the angle $(-\alpha)$ with each other but possess the same magnitudes as at A. Consequently $\langle\delta\vec{u}\times\delta\vec{h}\rangle$ vanishes, so that under the assumed conditions the dynamic demagnetizing field $\delta\vec{h}$ is "random in the average." Using (42), one obtains from (40)

$$\langle\vec{b}\rangle = \langle\vec{h}\rangle + i(4\pi M_s\gamma/\omega)\langle\vec{h}\rangle\times\langle\vec{u}\rangle, \qquad (43)$$

where the approximate equality is written as an equality for convenience. Now let $\vec{i}_z$ denote the direction, and M_z the magnitude, of the static magnetization at some arbitrary point. Then the equation

$$\langle\vec{u}\rangle = (M_z/M_s)\vec{i}_z \qquad (44)$$

is in accord with the fact that the static magnetization in a polycrystal represents the mean value (over V') of $\vec{M}_s$. Substituting (44) into (43), one finds upon com-paring the result with (24a) that μ, μ_t, μ_z, are given by the simple equations

$$\mu = 1, \qquad (45a)$$
$$\mu_t = -(4\pi M_s\gamma/\omega), \qquad (45b)$$
$$\mu_z = 1. \qquad (45c)$$

To take damping into account, one may add (31) (with $\vec{H}$ replaced by $\vec{H}'$) to the right hand side of (33). If one now introduces λ_r', and uses the assumption

$$\lambda_r' \equiv (4\pi\lambda_r/\omega) \ll 1, \qquad (46)$$

in addition to all those stated earlier, then one finds that μ_t is unchanged by the damping, but μ and μ_z take on the complex form (18a) and (18c), respectively. The resulting tensor components are determined by

$$\mu_1 = 1, \qquad (47a)$$
$$\mu_2 = \lambda_r', \qquad (47b)$$
$$\mu_{z1} = 1, \qquad (47c)$$
$$\mu_{z2} = \lambda_r'[1 - (M_z/M_s)^2], \qquad (47d)$$

where the fundamental questions connected with the (to the approximation considered here) damping term

(31) causes the applicability of (47) to be subject to the same limitations as the applicability of (32).

It should be emphasized that the derivation of (45) and (47) shows that these results are valid even if $\vec{i}_z$ and M_z vary from point to point. As stated in the discussion of Section IIH, such a variation could occur if, for example, the applied static field is nonuniform. However, under such conditions it is rather difficult to measure (or calculate) $\vec{i}_z$ and M_z, so that in practice (45) and (47) will probably be most useful when $\vec{i}_z$ and M_s are uniform. In that case (45),[3] and the prediction[3] of a Faraday rotation proportional to M_z throughout the hysteresis loop, have been verified experimentally by van Trier.[1]

3. *Arbitrarily magnetized ferrimagnetic media.* It is useful to begin this discussion with a brief review of some of the similarities and differences of ferrimagnetic and ferromagnetic media. In both types of media the elementary magnetic moment is primarily that associated with electron spin. The spins of certain electrons in these media are subject to a cooperative effect due to exchange forces (which may be represented by Weiss fields) and thus give rise to a spontaneous magnetization. The direction of the latter varies from domain to domain and is determined by the requirement that the direction-dependent part of the total free energy (due to magnetic interactions, magnetocrystalline anisotropy, magneto-elastic anisotropy, and exchange) be a minimum. However, the two types of media differ in the manner in which the relevant spins within a domain are oriented relative to each other. At the absolute zero of temperature, and apart from the quantummechanical zero-point fluctuations, the spontaneous magnetization in a ferromagnetic medium is due to parallel spins only, whereas the spontaneous magnetization of a ferrimagnetic medium is due to the fact that more spin magnetic moments have a component in one direction (that of the spontaneous magnetization $M_s\vec{u}$) than in the opposite direction.

In a ferromagnetic medium the spins usually wander about the crystalline lattice, so that it is only in the time average that a certain (noninteger) number of them are associated with each magnetic ion; in a ferrimagnetic medium, on the other hand, the spins may usually be regarded as staying on specific magnetic ions, the latter being arranged in an orderly manner on the crystalline lattice. The most common ferrimagnetic medium is a ferrite possessing the spinel structure. In this case all the ions (of whatever type) whose moments are parallel to $\vec{u}$ are located on one set of crystallographically equivalent lattice sites, thus constituting sublattice B, and all the ions (of whatever type) whose moments are antiparallel to $\vec{u}$ are located on another set of crystallographically equivalent lattice sites, thus constituting sublattice A. Néel showed that it is approximately correct to represent the exchange forces acting on the ions of one sublattice by means of a

negative Weiss field due to the ions of the other sublattice. For the present purposes one can, therefore neglect the so-called AA and BB interactions, and write

$$\vec{H}_{WB} = -\lambda_W(-M_{SA}\vec{u}), \qquad (48a)$$

$$\vec{H}_{WA} = -\lambda_W(M_{SB}\vec{u}), \qquad (48b)$$

where M_{SA}, M_{SB} and λ_W are positive numbers. The quantity λ_W is known as the molecular field coefficient. On this model, the magnitude of the spontaneous magnetization of the medium is given by the relation

$$M_s = M_{SB} - M_{SA}, \qquad (49)$$

in terms of the magnitudes of the spontaneous magnetizations of the sublattices.

It is physically plausible that a ferrimagnetic medium will act electromagnetically as if it were a ferromagnetic medium provided the following three conditions are satisfied:

1. The "net" Weiss field must be much larger than all other effective and real magnetic fields, so that one has

$$\lambda M_s \gg H_e. \qquad (50)$$

2. The frequencies under consideration must be smaller than those corresponding to the infared region, as required by assumption (e).

3. The g factor must be replaced by the effective g factor g_{eff}, as mentioned earlier.

Conditions 1 and 2 insure that the spin system will have sufficient "stiffness," at all frequencies under consideration, to maintain the antiparallel alignment of the sublattices A and B during the precessions. Thus the system is effectively a ferromagnet possessing a spontaneous magnetization M_s. Condition 3 means theoretically that γ must be replaced by Wangsness' expression[5] for γ_{eff}. In the case of two sublattices, γ_{eff} is given by

$$\gamma_{eff} = \frac{M_{SB} - M_{SA}}{(M_{SB}/\gamma_B) - (M_{SA}/\gamma_A)},$$

where γ_A and γ_B refer to the two sublattices. In practice however, the γ_{eff} that is used in the electromagnetic calculations can be regarded as an empirical constant to be taken from the results of resonance experiments.

By writing for each sublattice an equation of motion similar to (26), and using (48) and (49), Wangsness shows that under conditions 1 through 3 the results of Polder, (30), are indeed valid in a ferrimagnetic medium. Furthermore, Wangsness finds (in effect) that under these same conditions the writer's results, (45), are also valid in a ferrimagnetic medium.

It should be mentioned that Wangsness does not consider arbitrary magnetic interactions or a poly-crystalline structure in nonsaturated media, but he does consider geometrical demagnetizing factors in both saturated and nonsaturated media. The first point does not really constitute any loss of generality because the inequality (50) is assumed to hold. The second point simply means that some of the permeability tensor

rived by Wangsness relate $\vec{b}$ to the externally applied rather than to the actual $\vec{h}$ in the medium.

The equivalence of ferromagnetic and ferrimagnetic media, under the conditions 1 through 3, continues to old even if an arbitrary number of sublattices had to considered in a given situation. This assertion is oved theoretically by Wangsness. Furthermore, the perimental verification of (45) was carried out on a rrimagnetic rather than ferromagnetic medium.[1]

The electromagnetic behavior of a ferrimagnetic edium becomes more complicated (and the theory less raightforward) than indicated above as soon as the pproximation (50) cannot be made. This case occurs if e medium is near a "compensation point;" i.e., near a ate where the net magnetization or angular momentum nishes. Since only a few media possess compensation ints, Wangsness' treatment of such media will not be scussed here.

V. AN EQUATION THAT CONNECTS $\langle \vec{b} \rangle$ AND $\langle \vec{h} \rangle$ BY OTHER THAN A POINT RELATION

Throughout Section IV the assumption (i) will be scarded. It will be assumed, instead, that the relation tween $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ does contain spatial derivatives. hile no attempt will be made to specify the most neral form of this relation, it should be pointed out at the abandonment of assumption (i) necessarily eans that the $\langle \vec{b} \rangle$ vs $\langle \vec{h} \rangle$ relation is not representable a tensor.

The Exchange Effect and the Spin Wave Equation

A special physical model will be treated throughout ction IV. This model is the result of including the xchange stiffness" term

$$(2A/M_s{}^2)\vec{M} \times \nabla^2 \vec{M} \qquad (51)$$

(26). The term (51) is a classical torque density which presents the tendency of exchange forces to straighten t any nonuniform orientation of $\vec{M}$. Adding (51) to 6), one obtains the "spin wave equation"

$$(1/\gamma)\partial \vec{M}/\partial t = \vec{M} \times [\vec{H} + (2A/M_s{}^2)\nabla^2 \vec{M}], \qquad (52)$$

here $(2A/M_s{}^2)\nabla^2 \vec{M}$ may be regarded as an effective ld.

Before discussing the origin and consequences of 1), it is appropriate to recall that a nonuniform orienta- n of $\vec{M}$ exists inside the domain walls of any ferro- agnetic medium, so that (51) should be taken into count whenever the medium is not saturated. This n be done, for example, by adding (51) to (33). owever, as shown in Section IIIA, it is a consequence of mmetry and averaging considerations that in many ses the relation between $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$ is nevertheless ven by a tensor. In these cases the "exchange stiffness nstant" A, together with many other factors, deter- nes merely the values of the tensor components oc-

curring in (17). The absence of A in (45) is due to the fact that domain wall displacements are explicitly neglected in the special model of Section IIIB2.

Another situation in which a nonuniform orientation of $\vec{M}$ occurs is in the skin depth of a ferromagnetic metal. It is this situation that will be considered in the present section. Since the medium will be assumed to be saturated, (52) will be used, and no averaging of the resulting (nontensor type) relation between $\vec{b}$ and $\vec{h}$ will be necessary. In principle, however, an anisotropy torque could be incorporated into (52). The resulting relation between $\vec{b}$ and $\vec{h}$ could then be averaged in a polycrystalline medium to obtain a relation between $\langle \vec{b} \rangle$ and $\langle \vec{h} \rangle$. The latter relation would also be of the nontensor type because a large part of the effect of $\nabla^2 \vec{M}$ at a given point would depend on the distance of that point from the surface of the specimen; this part of the $\nabla^2 \vec{M}$ effect would evidently not be eliminated by the averaging process discussed in Section IIC.

The term (51) can be derived from the so-called Dirac "cosine coupling" between neighboring spins provided a number of specific assumptions are made. An alternative "derivation" of (51) is based on the simple assumption that a nonuniform orientation of $\vec{M}$ gives rise to an exchange torque (and exchange energy) depending upon the spatial derivatives of $\vec{M}$; if the orientation of $\vec{M}$ does not vary appreciably in an intera- tomic distance, then it immediately follows from sym- metry considerations that in a medium possessing cubic crystal structure the exchange torque must have the form (51). References to further discussions of these derivations are cited by Ament and Rado.[6]

B. The New Boundary Conditions and the Equivalent Isotropic Permeability

Some of the consequences of (52) for ferromagnetic resonance were first published by Kittel and Herring[17] but may have been obtained earlier by Macdonald.[18,19]. In the following, several aspects of Ament and Rado's[6] recent treatment will be outlined. Their work provides an explicit electromagnetic theory of exchange effects in ferromagnetic metals and satisfactorily accounts for the experimental results of Rado and Weertman.[20] The latter were the first to observe exchange effects in ferromagnetic resonance. Their experimental conditions were chosen on the basis of physical considerations and differed from those envisaged by the earlier theories.

To calculate a quantity which corresponds to a per- meability in the presence of exchange effects, the par-

[17] C. Kittel and C. Herring, "Effects of exchange interaction on ferromagnetic microwave resonance absorption," *Phys. Rev.*, vol. 77, pp. 725–726; March, 1950.

[18] J. R. Macdonald, Ph.D. thesis, Oxford; 1950.

[19] J. R. Macdonald, "Ferromagnetic resonance and the internal field in ferromagnetic materials," *Proc. Phys. Soc. Sec. A*, vol. 64, pp. 968–983; November, 1951.

[20] G. T. Rado and J. R. Weertman, "Observation of exchange interaction effects in ferromagnetics by spin wave resonance," *Phys. Rev.*, vol. 94, p. 1386; June, 1954.

ticular physical situation used in the experiments[20] will be treated. Consider a plane slab of ferromagnetic metal extending from the plane $y = o$ to a plane whose y-coordinate is much larger than the skin depth. Let the applied static magnetic field be along $\vec{i}_z$, and assume that the applied microwaves are plane waves normally incident (from air) upon the $y = o$ plane, the tangential component of their magnetic vector being along $\vec{i}_x$. Now combine (52) with the Maxwell equations (1) and (2), and note that the displacement current may be neglected because of the high conductivity of the metal. Using (27), and the linearizing assumptions (28), and putting $\vec{m}$ and $\vec{h}$ proportional to $\exp(i\omega t - ky)$, one then obtains a secular equation which is cubic in k^2. The solution of this equation yields three propagation constants (k_1, k_2, k_3) whose real part is positive; the corresponding three waves, propagating along $\vec{i}_y$, represent energy flow into the metal. Since in the no-exchange case there is only one such wave, the determination of the (complex) wave amplitudes requires, in this case of triple refraction, two new boundary conditions which must be a consequence of the exchange effect.

The exchange torques are due to forces that exist solely in the interior of the ferromagnetic metal. Consequently the total exchange torque per unit area of air-metal boundary must vanish. Thus one finds

$$\int_o^\infty \vec{M} \times (2A/M_s{}^2)\nabla^2\vec{M}\,dy$$
$$= (2A/M_s{}^2) \int_o^\infty (M_s\vec{i}_z + \vec{m})\times(\partial^2\vec{m}/dy^2)\,dy = 0. \quad (53)$$

Integrating by parts, and using the approximation (28b) as well as the requirement $A \neq 0$, one obtains from (53) the new boundary conditions

$$(\partial m_x/\partial y) = 0 \quad \text{at} \quad y = 0, \quad (54a)$$
$$(\partial m_y/\partial y) = 0 \quad \text{at} \quad y = 0. \quad (54b)$$

Eqs. (54) were first obtained by Macdonald[18] in his unpublished thesis, and were later derived independently by Ament and Rado.[6]

The standard Maxwellian boundary conditions, which require that h_x and e_x be continuous at $y = o$, may now be combined with (54) to obtain a system of four linear homogeneous equations that contain the field in air (h_{oz}) and the three fields in the metal (h_{1x}, h_{2x}, h_{3x}). From the condition that this system possess a non-vanishing solution, the value of the "surface impedance"

$$Z = (e_x/h_x)_{y=o} \quad (55)$$

may then be calculated. The result can be expressed in terms of an "equivalent isotropic permeability" which is defined by the equation

$$\mu_{equ} = \mu_1 - i\mu_2 = 2i(cZ/\omega\delta)^2, \quad (56)$$

where δ is the classical skin depth for permeability unity. Proceeding in this manner one obtains

$$\mu_{equ} = \frac{\eta - \Omega^2 + 2\psi(1+i)}{[\eta - \Omega^2 + \psi(1+i)]^2}, \quad (57)$$

where the parameters

$$\eta = H_o/(4\pi M_s), \quad (58a)$$
$$\Omega = \omega/(4\pi M_s\gamma), \quad (58b)$$
$$\psi^2 = A/(2\pi M_s{}^2\delta^2), \quad (58c)$$

are supposed to have such values that each of the quantities η, Ω^2, and ψ is negligible compared to unity.

The discussion[6, 20] of (57) and related results will be restricted to the following few comments:

1. If the damping term (31) had been included in (52), then (57) would have to be replaced by a more complicated expression[6] containing λ_r.

2. If exchange effects are negligible, then the μ_{equ} defined by (56) is simply given by the scalar permeability $\mu_x = b_x/h_x$. In the presence of exchange effects, on the other hand, μ_{equ} cannot be regarded as any kind of a (flux density/field) ratio. Thus μ_{equ} is, in this case, merely a measure of the surface impedance.

3. Eq. (57) shows that μ_2 does not vanish as long as δ is finite and A is not zero. The "magnetic" losses determining μ_2 are due to the combined effect of eddy current dissipation and exchange. These particular eddy current effects are not to be confused with the ordinary (non-magnetic) eddy current losses which exist even if A vanishes or is neglected.

4. Note that in the presence of exchange the value of μ_2 may become negative, as shown by (57) and by experiment.[20] This does not mean that the metal can generate microwaves. The possibility of a negative μ_2 merely demonstrates that in the presence of exchange the quantity μ_{equ} is not a $(\vec{b}/\vec{h})$ ratio so that there is no restriction on the sign of μ_2.

5. Eq. (57) can be used to obtain not only γ (and hence g) but also the important quantity A from experimental results.[20]

6. Eq. (57) remains valid, under certain conditions, even if the sample of ferromagnetic metal is not plane and the incidence of the microwaves is not normal.

In conclusion, the writer wishes to thank W. S. Ament for reading the manuscript and making useful comments.

Appendix I. The Conductivity Tensor

If the magnetoresistance and Hall effects are not negligible, then the conductivity can no longer be represented by the scalar σ used in (6) and (11). Instead, the relation between $\vec{E}$ and $\vec{J}$, in media in which the effects of crystallographic orientation are negligible, is given by the equations

$$E_x = \rho_\perp J_x - \rho_H J_y, \quad (59)$$
$$E_y = \rho_H J_x + \rho_\perp J_y, \quad (60)$$
$$E_z = \rho_\parallel J_z, \quad (61)$$

where the rectangular co-ordinate system is defined by the fact that the positive z-axis is parallel to the static magnetization or static magnetic field. The tensor

omponents $\rho_\perp$ and $\rho_{||}$ denote the resistivity measured rpendicular and parallel to the z direction, respectively, nd ρ_H denotes the Hall resistivity. The quantities $_\perp$, $\rho_{||}$, ρ_H may depend on the static part of the magetization and/or magnetic field. Relations analogous to 59), (60), (61) evidently apply to the average vectors $\vec{E}$ and $\langle\vec{J}\rangle$ also, provided the assumptions discussed Section II are fulfilled. Finally, (59), (60), (61), apply qually well to the time dependent components $\vec{e}$ and $\vec{j}$ $\langle\vec{e}\rangle$ and $\langle\vec{j}\rangle$.

A simple calculation based on (59), (60), (61) yields he components of the conductivity tensor. The latter ay then be used to obtain a tensor form of the effective" permittivity. Consequently one may expect 59), (60), (61) to give rise to electromagnetic phenom- na analogous to those caused by the tensor permeability. n particular, one can calculate a "Hall rotation" of nearly polarized plane waves in a ferromagnetic med- um. If at microwave frequencies the extraordinary Hall ffect in ferromagnetic media is much larger than the rdinary Hall effect, as it is at zero frequency, then he writer's preliminary calculation shows that under ertain conditions the Hall rotation, like the analogous henomenon of Faraday rotation, should be proportional the static magnetization. However, reliable estimates f the magnitude of the Hall rotation cannot now be ade because measured values of ρ_H at microwave requencies do not seem to exist.

Appendix II. Calculation of Losses

Rewriting the Maxwell equations (1) and (2) for the ime dependent vectors, and abbreviating $\partial/\partial t$ by a dot," one obtains

$$\vec{\nabla}\times\vec{e} = -(1/c)\dot{\vec{b}},$$

$$\vec{\nabla}\times\vec{h} = (4\pi/c)\vec{j} + (1/c)\dot{\vec{d}},$$

o that the use of the vector identity

$$\vec{h}\cdot\vec{\nabla}\times\vec{e} - \vec{e}\cdot\vec{\nabla}\times\vec{h} = \vec{\nabla}\cdot(\vec{e}\times\vec{h}),$$

ombined with the divergence theorem, leads to the well nown Poynting relation

$$(1/4\pi)\int\vec{h}\cdot\dot{\vec{b}}\,dv + (1/4\pi)\int\vec{e}\cdot\dot{\vec{d}}\,dv + \int\vec{j}\cdot\vec{e}\,dv = -(c/4\pi)\int\vec{e}\times\vec{h}\cdot\vec{n}\,da, \quad (62)$$

vhere the integrals on the left and right hand sides of 62) are to be taken over the volume and surface, re- pectively, of a given sample, and $\vec{n}$ denotes a unit ector normal to the surface. Eq. (62) expresses the onservation of energy in adiabatic situations, and the onservation of Helmholtz free energy in (the more requently encountered) isothermal situations. For the urposes of the present paper, the significance of (62) s that the time average of the three terms on the left and side represents the energy per second dissipated in he sample due to magnetic, dielectric, and conduction Joule) losses, respectively, and that the time average

of the right hand side represents the net inflow, per second, of electromagnetic energy into the sample.

The question arises whether the losses can be calcu- lated if the average rather than the ordinary vectors are used. On the basis of the arguments given below, it is now suggested that (62) is rigorously valid in form, and its left-hand side approximately valid in interpreta- tion, if the ordinary vectors are replaced by the average vectors. To show this, define $\vec{\delta h}$ by

$$\vec{h} = \langle\vec{h}\rangle + \vec{\delta h}, \quad (63)$$

with similar relations defining $\vec{\delta b}$, $\vec{\delta e}$, $\vec{\delta d}$, $\vec{\delta j}$. According to Section IIC, the average $\langle\ \rangle$ is taken over a volume V' which is much larger than the inhomogeneities. Consequently the "cross terms" $\int\langle\vec{h}\rangle\cdot\vec{\delta b}\,dv$ and $\int\vec{\delta h}\cdot\langle\vec{b}\rangle\,dv$ both vanish, and one obtains

$$\int\vec{h}\cdot\vec{b}\,dv = \int(\langle\vec{h}\rangle\cdot\langle\vec{b}\rangle + \vec{\delta h}\cdot\vec{\delta b})\,dv.$$

Since similar arguments apply to the other terms of (62), this equation becomes

$$\int[(1/4\pi)\vec{h}\rangle\cdot\langle\dot{\vec{b}}\rangle + (1/4\pi)\langle\vec{e}\rangle\cdot\langle\dot{\vec{d}}\rangle + \langle\vec{j}\rangle\cdot\langle\vec{e}\rangle]\,dv$$

$$+ \int[(1/4\pi)\vec{\delta h}\cdot\dot{\vec{\delta b}} + (1/4\pi)\vec{\delta e}\cdot\dot{\vec{\delta d}} + \vec{\delta j}\cdot\vec{\delta e}]\,dv$$

$$= \int(-c/4\pi)\langle\vec{e}\rangle\times\langle\vec{h}\rangle\cdot\vec{n}\,da + \int(-c/4\pi)\vec{\delta e}\times\vec{\delta h}\cdot\vec{n}\,da, \quad (64)$$

or, in abbreviated form,

$$L + \Delta L = R + \Delta R, \quad (65)$$

where L denotes the first integral in (64), etc.

Rewriting the "average" Maxwell equations (8) and (9) for the time dependent vectors, one obtains

$$L = R, \quad (66)$$

or in unabbreviated form,

$$(1/4\pi)\int\langle\vec{h}\rangle\cdot\langle\dot{\vec{b}}\rangle\,dv + (1/4\pi)\int\langle\vec{e}\rangle\cdot\langle\dot{\vec{d}}\rangle\,dv$$

$$+ \int\langle\vec{j}\rangle\cdot\langle\vec{e}\rangle\,dv = -(c/4\pi)\int\langle\vec{e}\rangle\times\langle\vec{h}\rangle\cdot\vec{n}\,da. \quad (67)$$

Although Eq. (67) is rigorously correct, it should be emphasized that its left-hand side does not express correctly the energy per second dissipated in the sample.

Note, however, that the time average of ΔR evidently represents the additional scattering due to the inhomo- geneities, i.e., the effect of the inhomogeneities on the net rate of inflow of electromagnetic energy into the sample. Since the linear dimensions of V' are small compared to $\lambda/4$ (see Section IIC), it follows from physical considera- tions (Raleigh scattering) that

$$\Delta R \ll R \quad (68)$$

is a fair approximation. In view of (68), it follows from (66) and (65) that $\Delta L \ll L$ is also a fair approximation. Consequently the left-hand side of (67) may be inter- preted, to a fair approximation, as expressing the energy per second dissipated in the sample. Eq. (67) was dis- cussed in Section IID, example 4, and used in Section IIIA, (20) and (21a).

Plasma Oscillations

D. GABOR†

Summary—This paper is a report on the investigations by the author and collaborators F. Berz, E. A. Ash, and D. Dracott at Imperial College. F. Berz has theoretically investigated wave propagation in a uniform plasma and found that even in the absence of collisions only damped waves can arise, because the fluctuating velocity distribution contains a term, overlooked by previous authors, which represents a flowing-apart of the electron density. The cutoff due to this effect alone is at about 1.15 of the Langmuir frequency, and the shortest wavelength at about 20 Debye lengths.

Experimental investigations by E. A. Ash and D. Dracott extending over 5 years have at last elucidated the paradox of the existence of Maxwellian electron distributions in the positive column of arcs at low pressures. The interaction is not between electrons and electrons but between these and an oscillating boundary sheath. The sheath was explored by an electron beam probe and oscillations of about 100 mc observed under conditions when the plasma frequency in the arc was about 500 mc. Electrons diving into the boundary sheath spend about one cycle in it, during which time they can gain or lose energies of the order of several volts. Possible applications to radio astronomy are briefly suggested.

INTRODUCTION

THIS IS a report on the investigations, theoretical and experimental, carried out at the Imperial College, London, with my students and collaborators Dr. F. Berz, Dr. E. A. Ash and Mr. D. Dracott, in the last five years, and which are still mostly unpublished. Our investigations were stimulated by a desire to understand the phenomenon which I propose to call "Langmuir's Paradox," discovered by Irving Langmuir 30 years ago, and which so far has defied explanation. After five years' work we succeeded in reducing it to a hitherto unsuspected type of plasma-boundary oscillation, with which I will deal in the second part of the paper. In order to understand plasma-boundary phenomena it is useful to study first the uniform plasma, and I propose to start with a report on this side of our investigations.

THE PLASMA AS A WAVE-PROPAGATING MEDIUM

I. Langmuir and L. Tonks,[1,2] were the first to find, in 1928, that an idealized plasma without thermal agitation, with n electrons and an equal number of ions per unit volume is capable of longitudinal oscillations with a radian frequency

$$\omega_o{}^2 = 4\pi e^2 n/m. \tag{1}$$

This was the discovery of plasma oscillations, but not yet of plasma waves, because waves can carry energy only if the frequency depends on the wavelength, while Langmuir's formula gives a group velocity *nil*. Soon afterwards J. J. Thomson[3] was able to prove what

Langmuir and Tonks suspected, that in the presence of thermal agitation of the electrons the frequency will depend on the wavelength, and that (1) gives in fact the limit for infinite waves. J. J. Thomson gave the first dispersion law of plasma waves, which subsequent investigation showed to be not quite accurate.

There the matter rested for a long time, until the important paper in 1945 by A. Vlassov,[4] which E. Landau[5] in 1946 subjected to penetrating and illuminating criticism, followed, a few years later by the thorough study of D. Bohm and E. P. Gross.[6] Starting from an integral equation, due to Vlassov, which they derived by a novel method, these authors established the dispersion law of plasma waves shown in Fig. 1. These are supposed to be undamped, progressive waves; long waves are damped only by collisions, though Bohm and Gross, as Langmuir before them, adduced good reasons to believe that there would be strong damping for waves approaching the Debye length,

$$\lambda_D = (KT/4\pi e^2 n)^{1/2} \tag{2}$$
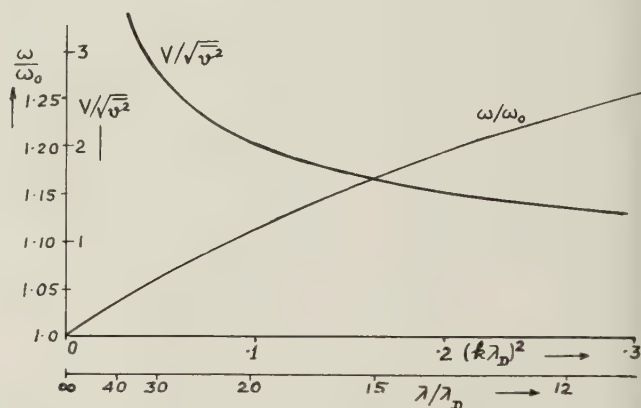
Fig. 1—Dispersion law of plasma waves according to Bohm and Gross, 1949.

We were somewhat concerned about the fact that the integral equation from which the dispersion law is derived contains an improper integral, and the procedure of Bohm and Gross is certainly not justifiable on mathematical grounds. On the other hand the result appears physically reasonable, and F. Berz and I tried to justify it by different mathematical procedures. I will now follow F. Berz's approach,[7] as this is by far the more rigorous, omitting the recondite mathematics, which will be published elsewhere.

---

† Imperial College, London, England.
[1] I. Langmuir, "Oscillations in ionized gases," *Proc. Nat. Acad. Sci. Amer.*, vol. 14, pp. 628–000; 1928.
[2] L. Tonks and I. Langmuir, "Oscillations in ionized gases," *Phys. Rev.*, vol. 33, pp. 195–210; 1929.
[3] J. J. Thomson and G. P. Thomson, "Conduction of Electricity in Gases," vol. 2, Cambridge University Press, Cambridge, p. 353; 1933.

[4] A. Vlassov, "On the kinetic theory of an assembly of particles with collective interraction," *Jour. Phys.*, *USSR*, vol. 9, pp. 25–40; 1945.
[5] E. Landau, "On the vibration of the electronic plasma," *Jour. Phys. USSR*, vol. 10, pp. 25–35; 1946.
[6] D. Bohm and E. P. Gross, "Theory of plasma oscillations," *Phys. Rev.*, vol. 75, pp. 1851–1864; 1949; vol. 75, pp. 1864–1876; 1949; vol. 79, pp. 992–1001; 1955.
[7] F. Berz, "Plasma Electron Oscillations," Thesis, London; 1955.

Berz found that Vlassov's integral equation has in fact no rigorous solution corresponding to undamped progressive waves. On the other hand if one assumes a damping; e.g., in time, however small, one can construct within limits waves with *any* wave number $k = 2\pi/\lambda$ and *any* ω (with a small imaginary component), if only one assumes the right sort of velocity distribution of the electrons. This is obtained by adding to the fluctuating component of the distribution assumed by Vlassov and by Bohm and Gross a term of the form

$$\frac{CE}{v_x - \omega/k} \exp[ik(x - v_x t)], \qquad (3)$$

where E is the electric field amplitude in the wave, v_x the x-component of the electron velocities, and C an *arbitrary* constant. By choosing C appropriately one can produce any slightly damped wave, hence *there is no definite dispersion relation*. (This has been pointed out before by Landau, in his criticism of Vlassov, though by a different argument.) On the other hand, if the wave is to be undamped, this distribution function has a singularity at $v_x = \omega/k$, that is to say at the wave velocity V, as indicated in Fig. 2. This is a nonphysical solution, and the only choice in this case is $C = 0$, which gives no waves at all!
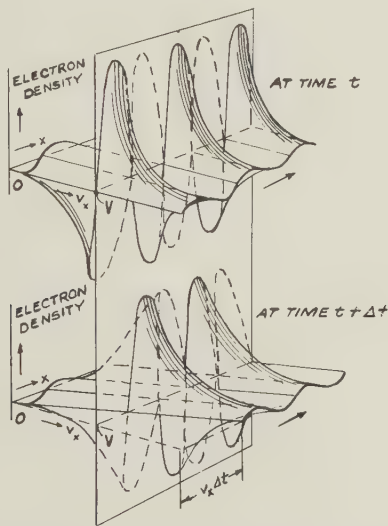
Fig. 2—Disturbance in the distribution function required to maintain undamped plasma waves according to F. Berz, 1955.

Eq. (3) represents a velocity distribution which is carried away with the electron velocities v_x, and is therefore dispersed in space. This represents a source of damping; the electron density waves do not keep together as previous authors assumed, but flow apart.

On the other hand as regards damped waves, the remarkable fact emerges that there exists a class of *waves with minimum damping*, with a definite dispersion relation, which for long waves, where the damping is small, approaches the law of Bohm and Gross. F. Berz has worked this out with great mathematical and numerical labor for a plasma with Maxwellian electron-velocity distribution, and obtained the results shown in Fig. 3.

(These relate to waves damped in time; i.e., with a complex ω, not in space, but it is very likely that they apply also approximately for complex k.) It is seen that the damping, which is very small up to a frequency $1.08\omega_o$ becomes so strong at about $1.15\omega_o$ that one can say that above this frequency plasma waves do not exist. The frequency $1.15\omega_o$ corresponds to a wavelength of about 20 Debye lengths; hence the cut-off is at much longer waves than hitherto believed.[1,8,9]

P_e : *NUMBER OF PERIODS IN WHICH INTENSITY DROPS TO $1/e$*
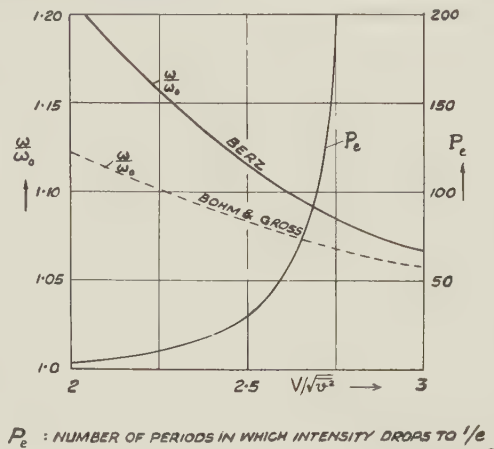
Fig. 3—Damping of plasma waves (without collisions) according to F. Berz, 1955.

Berz has also investigated the effect of collisions between electrons and neutral atoms on the damping. The result is, roughly speaking, that the plasma oscillations will survive only for the interval between two collisions. This restricts the region in which plasma waves can exist at all, in the conditions usually prevailing in gas discharges, to pressures below $10^{-3} - 1$ mm, according to the nature of the gas. It still leaves a wide field for astrophysical phenomena in which plasma waves may be of importance.

SELF-MAINTAINING PLASMA-BOUNDARY OSCILLATIONS

What I propose to call "Langmuir's Paradox" is the existence of a Maxwellian distribution of electron energies in the positive column of low-pressure gas discharges, or for that matter the existence of the positive column itself. Neither has a right to exist, according to the present theory. The Maxwellian energy distribution of the electrons, discovered over 30 years ago by I. Langmuir and H. Mott-Smith,[10] appears to indicate a very intense interchange of energy between the electrons, for which the existing theories of Coulomb interaction fail to account. The discrepancy in the case of very low pressure mercury arcs may be as much as 3 or 4 orders of mag-

[8] D. Pines and D. Bohm, "A collective description of electron interactions, III, Coulomb interactions in a degenerate electron gas," *Phys. Rev.*, vol. 92, pp. 609–625; 1953.

[9] D. Gabor, "Wave theory of plasmas," *Proc. Roy. Soc.* A, vol. 213, pp. 73–86; 1952.

[10] I. Langmuir and H. Mott-Smith, "Studies of electric discharges in gases at low pressure," *Gen. Elec. Rev.*, vol. 27, pp. 449, 538, 616, 762; 1924.

nitude. As the theory might well have been at fault, E. A. Ash and the author[11] have measured the interaction of electrons with thermal plasmas and with electron clouds directly, but found that the somewhat crude theory covers the facts surprisingly well. One can say, without much exaggeration, that in a mercury arc of a few microns pressure an electron will move through the plasma of the positive column, between two collisions with the negative boundary sheath which covers the walls, practically as through a vacuum; it will not exchange any noticeable energy with other electrons. How then is it possible that the "Maxwellian tail" of fast electrons is constantly replenished, in spite of the walls always swallowing up the fastest electrons?

It has been the fashion, for quite a time, to ascribe everything which is inexplicable in gas discharges to plasma oscillations, but E. A. Ash[12] soon proved that an electron beam fired across the plasma of a mercury arc was diffused only about as much as expected on the basis of the interaction theory, that is to say hardly at all. If oscillations were at the root of Langmuir's paradox, there was only one place where they could hide; in the sheath at the wall of the discharge tube. And this is where we found them in the end, by means of the apparatus shown schematically in Fig. 4.
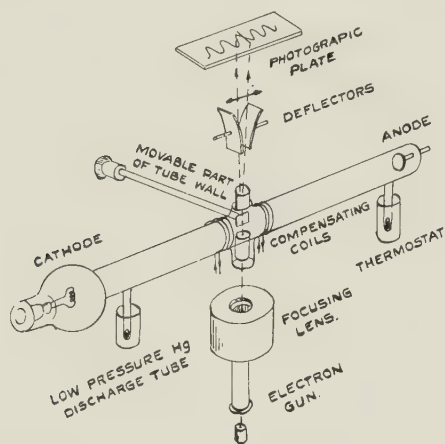
Fig. 4—Apparatus for observing plasma-boundary oscillations according to Ash, Dracott, and Gabor.

A dc mercury arc was maintained in a long discharge tube of about 3-cm diameter, with carefully controlled mercury pressure, which was varied between 0.5 and 3 microns in the course of the experiments. The electric field in the positive column region is explored by a fine electron beam, of only 0.1 mm diameter, which is fired across the arc, at a few millimeters distance from the tube wall. A part of the wall, a plane plate, 2 cm square was made movable by a micrometer screw. In some experiments this plate was made of glass, in others of

metal. If this is negative with respect to the arc plasma it is covered by the negative sheath, well known from the probe theory of Langmuir and Mott-Smith, which in our experiments was 1 − 2 mm thick.

The experiment was first carried out by Ash in 1952 with 1,000 ev exploring electrons, and it appeared to confirm our expectations. The electron beam was deflected as a whole if it traversed the sheath region, nearly by the amount as expected from Langmuir's sheath theory; it did not form a spot on the fluorescent screen but a line, at right angles to the tube wall. This could hardly be interpreted otherwise than revealing the presence of an oscillating electric field in the sheath, yet it took two more year's work by Mr. D. Dracott until we had complete certainty.

We made sure in the early experiments, that the line was not due to a diffusion effect, by cutting down the beam diameter to 1/40 of a mm, and also that we were not producing the oscillations ourselves, by varying the intensity of the exploring beam within 3 orders of magnitude, from 0.01 to 10 microamperes. Neither had the slightest influence on the length of the line, which thus had to be attributed to a spontaneous oscillating field. But all attempts to measure the frequency of these oscillations by catching the beam electrons in collectors and amplifying the ac current failed completely, because of the difficulties of matching the beam to vhf amplifiers. In the end we succeeded by direct oscillography. In these experiments the beam voltage had to be raised to 20,000 volts, and the electron gun cathode heated to the limit, as the beam had to be kept to the very small diameter of 0.1 mm and thus the usual resources of electron optics, as utilized in modern CR oscillographs, were not at our disposal.

The results are shown in Fig. 5. The first two oscillograms are taken with a long time scale; their length is about one millisecond. What one sees here is the irregular envelope of the vhf electronic oscillations. In order to see these clearly the sweep speed had to be increased until the duration recorded was of the order of 30 millimicroseconds only, as in the last two oscillograms. These revealed at last the oscillations, which were found to be not quite sinusoidal, with a fundamental frequency of about 100 mc. The frequency was not quite constant, it varied in oscillograms taken in quick succession within about 20 per cent, but this was much less than the variation in amplitude, shown in the first two oscillograms. It may be noted that the frequency of the plasma-boundary oscillations was 100–120 mc when the plasma frequency calculated from 1. with n substituted from probe measurements was about 500 mc. This corresponds to $n = 3.10^9/\text{cm}^3$.

Having now made sure of the character of the phenomenon, we can with confidence interpret the observations of Ash and Dracott on the details of the field distribution inside the oscillating sheath. The electron probe allows us for the first time to check the sheath

[11] E. A. Ash and D. Gabor, "Experimental investigations on electron interaction," *Proc. Roy. Soc.*, A, vol. 228, pp. 447–490; 1955.

[12] E. A. Ash, "Electron Interaction," Thesis, London; 1952.
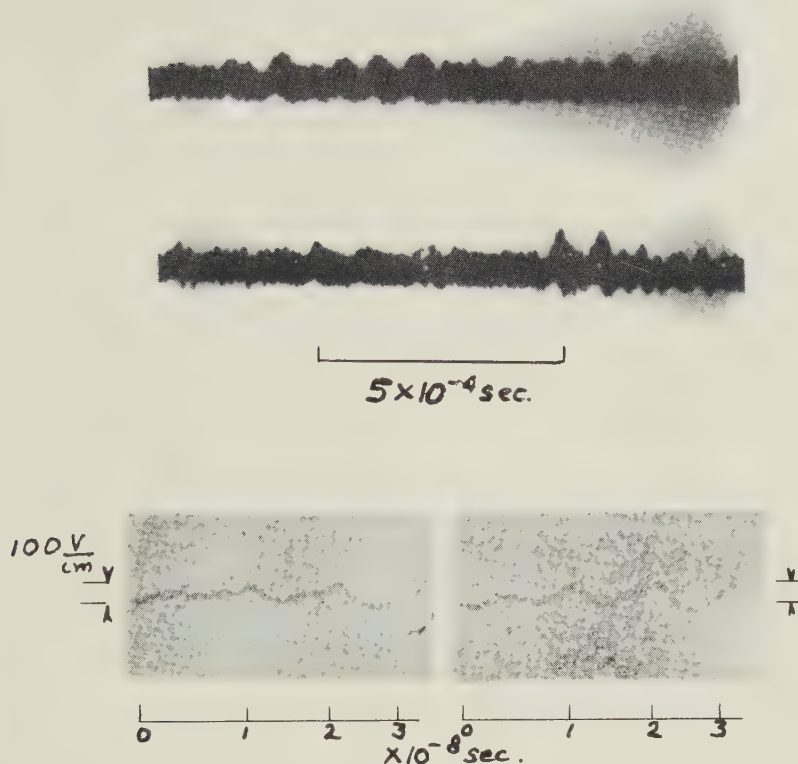
Fig. 5—Oscillations of plasma-boundary oscillations in a low-pressure Hg arc. (a) Two records of boundary oscillations on a long time scale, showing the irregular character of the envelope. (b) Two records of boundary oscillations on a very short time scale, showing the basic frequency of about 100 mc.

theory of Langmuir and Mott-Smith. It is well known that the sheath cannot be explored by a second Langmuir probe, as this immediately surrounds itself with a sheath of its own. But the thin electron probe allows measuring both the stationary and the oscillatings parts, E_o and E_n of the electric field intensity. The dc component is measured by compensating the mean deflection by means of the weak magnetic coils shown in Fig. 3. This gives E_o, and from this by integration one obtains V_o, the mean potential in the sheath. The results for one typical case are shown in Fig. 6. (Mercury saturation temperature 10 degrees C, pressure 0.5 micron, arc current 300 ma, $n = 10^9$, electron temperature $T -$ 6,000 degrees.)

The steady part of the field is nearly if not exactly in agreement with the Langmuir-Mott-Smith sheath theory. The fact that the field at the plasma edge does not fall to zero but only to a small constant value agrees, at least qualitatively, with the more exact Tonks-Langmuir theory of the positive column.

The oscillating component is of course quite unforeseen by previous theories. Its intensity is remarkably high; the peak-to-peak value of the field intensity shown in Fig. 6 exceeds the mean field over most of the sheath; outside $x = 1.2$ it is so strong that at one phase the electrons are actually dragged into the boundary sheath instead of being repelled by it!

It might appear surprising that an effect of this magnitude has not been discovered before, but it is in fact a very elusive effect. Like many other people we
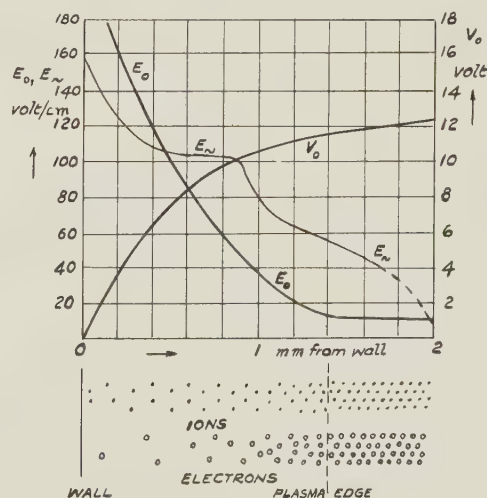
Fig. 6.

have tried to detect oscillations by attaching amplifiers to probes, but without any success.[13] The explanation is that the oscillations maintain themselves with practically

[13] Strong noise can be detected of course if a magnetic field of sufficient intensity is produced parallel to the wall, but this is quite another phenomenon. In this case the sheath changes into a sort of "preoscillating magnetron," with the plasma as cathode.

no supply of energy. They do not radiate outwards, because longitudinal oscillations do not propagate through the glass, nor do they propagate into the plasma, because the frequency of the boundary oscillations is well below the plasma cutoff frequency. If one tries to extract energy from them, the oscillations collapse immediately.

Yet the oscillations have a very important effect which has been known for a long time, and this is the phenomenon of "electron temperatures," in low pressure gas discharges, first discovered over thirty years ago. The oscillating boundary sheath is a *very efficient exchanger* of electron energies, because the electrons spend very nearly a full cycle in it. This is easy to see, if one approximates the part of the $V_o(x)$ curve near the sheath edge, which contains the great majority of electrons, by a parabola. It is well known that electrons in a parabolic potential trough oscillate with a constant frequency, irrespective of their energy, and in the case shown in Fig. 6 this works out to about 60 mc. Hence during the time in which an electron dives into the sheath and returns again into the plasma, that is to say performs a half-cycle of oscillation in the parabolic trough, the oscillating field of 100–120 mc goes very nearly through a full cycle. It can thus happen that one electron which dives into the sheath gains energy both on its outward and on its return journey, while an electron half a cycle later will lose energy both ways. As the oscillating field is of the order 50 v/cm, and the electrons move in it for a length of the order 1 mm, they will thus exchange several ev of energy at every dip. It is easy to show that the mixing-up mechanism is so efficient, that even electrons of uniform velocity dipping into the sheath at random time will acquire a velocity distribution which cannot easily be distinguished from a Maxwellian. Electrons which have a certain energy scatter to start with will approach the Maxwellian distribution even better.

The purpose of the oscillations is easy to see; it makes it possible for the arc *to survive* in conditions when otherwise it could not exist by producing at small cost the small number of electrons with energies sufficient for ionization, to replace those lost to the wall. If one has worked with gas discharges, one can hardly help thinking of the plasma as something "viable," something which fills out as much as possible of the discharge space and uses surprizing tricks for adapting itself to adverse conditions, but this, I think, is the most surprizing of all its tricks. There is good reason to believe that all low pressure, low current arcs maintain themselves by

plasma-boundary oscillations. In mercury the collision mechanism takes over probably only near 100 micron pressure. We could not detect the oscillations at the highest pressures we have used; they appeared to fade out as good as completely at 3 microns, but the reason is in all probability, that though they existed, they were no longer coherent over the length of 1 cm in which we explored the sheath.

The mechanism by which the oscillations maintain themselves is also easy to understand in principle, but not so easy to explain in detail. In principle this is evidently a sort of reflex-klystron, in which the electrons lose a little more energy on the whole than they gain, the small difference serves to cover the losses. These arise by a slight penetration of the field into the plasma or by the interaction of sheath areas which swing in antiphase. What is not so easy to understand, is how this oscillation can maintain itself with an electron stream which is so nearly Maxwellian. It cannot of course be completely Maxwellian, as this would correspond to thermal equilibrium, at which no organized oscillations are possible. We believe that the oscillation maintains itself on a slight deficiency in slow electrons compared with a perfect temperature distribution. The high energy "tail" is, as we have checked by numerous probe measurements, so nearly Maxwellian that it cannot account for the vigorous oscillations observed.

Though we have not yet succeeded in obtaining a complete theory of this complicated and highly non-linear phenomenon, we have got sufficiently far in understanding it for venturing a rather general conclusion:—*Plasma-boundary oscillations are likely to arise at almost any type of boundary if the plasma is off thermal equilibrium.* If there are boundaries of a type which do not absorb longitudinal waves, these oscillations constitute a powerful *mixing mechanism* for increasing the entropy, additional to collisions.

It is possible and even likely that this phenomenon plays a certain part in the generation of cosmic radio noise. Dust particles might constitute the boundaries and even in their absence the spontaneous density fluctuations might constitute fluid and transient boundaries, sufficient for producing oscillations, negligible on a laboratory scale but possibly important on a cosmic scale. Boundary oscillations on dust particles or around transient density-nuclei will, in general, possess an oscillating dipole moment so that a certain small amount of radiation can arise. I propose to report on this elsewhere.

Theory of Ferrites in Rectangular Waveguides[*]

K. J. BUTTON AND B. LAX[†]

Summary—Reciprocal and nonreciprocal propagation of electromagnetic energy in an infinitely long rectangular waveguide partially filled with one or two ferrite slabs is described. Methods for obtaining exact solutions of the transcendental equations usually encountered in these boundary value problems are demonstrated for several structures. Calculations are carried out for a lossless ferrite and the phase constant is plotted as a function of the ferrite slab thickness. The cutoff conditions for the lowest TE mode are evaluated in terms of the ferrite slab thickness. New modes, not associated with the empty waveguide modes, are analyzed as ferrite dielectric modes, their propagation characteristics are discussed and the rf electric and magnetic field patterns are plotted. The rf electric fields are plotted for all reciprocal and nonreciprocal modes and the appropriate field configurations are used to explain the operation of ferrite cutoff isolators, the field-displacement isolator, the field-displacement circulator, and the nonreciprocal phase shifter. Solutions above ferromagnetic resonance are shown and the *E*-fields are plotted. A brief comparison of the operation of dispersive devices at high and low frequencies is made. The calculations are extended to include absorption loss, and nonreciprocal attenuation is plotted as a function of slab position near resonance.

INTRODUCTION

ELECTROMAGNETIC problems involving ferrites in waveguides, lead, with a few exceptions, to transcendental equations which are rather complicated. In general it is difficult to evaluate the propagation constant from these equations. The situation is especially complex when one considers a waveguide partially filled with ferrite material. In particular, the problem of the ferrite rod in circular waveguide, which is of practical importance, has been extremely difficult to analyze. The only explicit expressions that have been obtained in a problem of this type are the approximations using perturbation theory, or the equivalent, for a very thin ferrite sliver in circular guide. However, there is one class of problems that can be evaluated exactly. This is the rectangular waveguide partially filled with one or two ferrite slabs and propagating in TE modes which have an electric field component only in the direction of the applied dc magnetic field. A typical structure which yields to exact analysis is shown in Fig. 1. The discussion in this paper will be devoted to an outline of a method for obtaining solutions and their applications. A description of the behavior and configuration of a number of modes for a selected range of the physical parameters involved, namely, the slab thickness, the internal dc field, the ferrite magnetization, the operating frequency and the waveguide width will be considered. The results will be used to explain the operation of several rectangular waveguide devices.
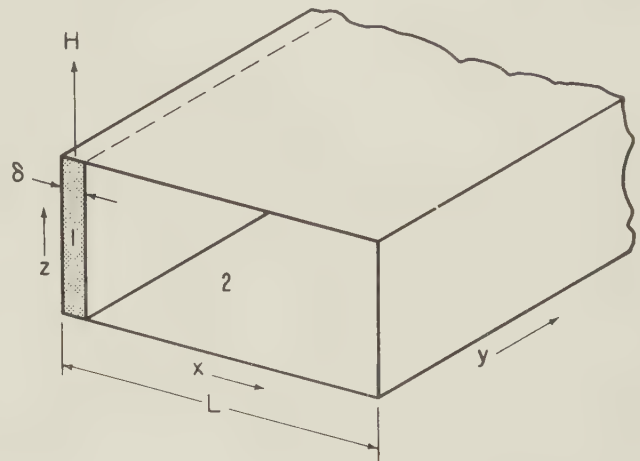
Fig. 1—A single ferrite slab against the side wall of a rectangular waveguide of internal width L.

ANALYSIS OF THE SINGLE-SLAB STRUCTURE

Maxwell's Equations

It is assumed that the ferrite is magnetically saturated by the applied dc field and that the region of operation is either below or above ferromagnetic resonance. Therefore the losses are assumed to be negligibly small so that the complex propagation constant contains only the phase term β. The validity of this assumption will depend, in some cases, upon the material having a relatively narrow resonance line width.

In a nonconducting ferromagnetic medium, Maxwell's equations are

$$\nabla x E = -j\omega\mu_o(1 + \overset{\leftrightarrow}{x}) \cdot \vec{h} \qquad (1)$$
$$\nabla x \vec{h} = j\omega\epsilon\vec{E},$$

where the time dependence $\exp(j\omega t)$ is to be assumed throughout. The wave equation expressed in terms of the rf magnetic field is then

$$\nabla x \nabla x \vec{h} = \omega^2\epsilon\mu_o(1 + \overset{\leftrightarrow}{x}) \cdot \vec{h}.$$

If the medium is magnetized in the z direction, $\overset{\leftrightarrow}{x}$ denotes the magnetic susceptibility tensor:

$$\begin{pmatrix} x_{xx} & x_{xy} & 0 \\ -x_{xy} & x_{xx} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

For these transverse electric modes, the components of $\vec{h}$ are assumed to have the form $\exp(jk_m x)\exp(-j\beta y)$ within the ferrite medium, where k_m is obtained from

$$k_m^2 = \omega^2\epsilon\mu_{eff} - \beta^2 \qquad (2)$$

and

$$\mu_{eff} = \mu_o\left[\frac{(1 + x_{zz})^2 + x_{zy}^2}{1 + x_{zz}}\right].$$

A similar procedure applied in the empty region of the guide yields the familiar relation

$$k_a{}^2 = \omega^2 \epsilon_o \mu_o - \beta^2. \tag{3}$$

Whenever k_m or k_a are imaginary, for convenience we shall substitute $K_m = -jk_m$ or $K_a = -jk_a$.

The Single-Slab Structure

It is now possible to consider the solution of an actual case such as that of the transverse electric modes in a rectangular waveguide containing a vertically magnetized ferrite slab against one side wall as shown in Fig. 1. Since the transverse electric field intensity must vanish at the side walls, either the circular or hyperbolic sine function must be used to describe the E-fields in each region. For real k_a and k_m,

$$E_z = [E_1 \sin k_m x] \exp(-j\beta y) \qquad \text{(Region 1)} \tag{4}$$
$$E_z = [E_o \sin k_a(L - x)] \exp(-j\beta y) \qquad \text{(Region 2)},$$

where L is the width of the waveguide. The expressions for the rf magnetic fields may be obtained from Maxwell's equations through the use of (4).

The boundary conditions at the face of the ferrite require the continuity of the tangential components of $\vec{E}$ and $\vec{h}$. These two boundary conditions result in two homogeneous linear equations in E_o and E_1 whose determinant must vanish. The resultant transcendental equation cannot be solved explicitly for the phase constant, β. However, it can be solved for the waveguide width L, as follows:

$$L = \delta + \frac{1}{k_a} \tan^{-1}\left(\frac{\dfrac{-k_a \mu_{eff}}{\mu_o}}{\dfrac{j\beta}{\theta} + k_m \cot k_m \delta} \right), \tag{5}$$

where $\theta = (1 + \chi_{xx})/\chi_{xy}$. The appearance of the linear term in β indicates that propagation associated with this particular structure is nonreciprocal. For a given slab thickness δ, the choice of a suitable trial value of β permits the evaluation of the waveguide width L.

Numerical Solution

The above procedure was carried out for a microwave frequency of 9,000 mc, a standard waveguide width of 2.286 cm, a saturation magnetization of 3,000 gauss, a dielectric constant of 10 and an internal dc field[1] of 1,000 oersteds. Several values of δ were used with appropriate choices of β and the result is shown in Fig. 2. Curves A and B show the phase constant as a function of slab width for forward and reverse propagation in the lowest TE mode corresponding to the fundamental waveguide mode. The electric field configurations shown have been computed from (4). It is clear that the electric fields have been slightly distorted from the empty waveguide pattern in such a way that the mode is required to propagate in an effectively narrower waveguide. As the

---

[1] The calculations are plotted in terms of the actual internal dc field that is found within the ferrite, namely, $H_i = H$ (applied)$-N_z M$. N_z is the usual demagnetization factor which depends only upon the geometry of the sample.

slab thickness is increased, cutoff is reached. For curve A and B, $\beta = 0$ at a thickness of about $4\frac{1}{2}$ mm. Under these conditions, this point represents the limiting thickness of ferrite which may be used for a nonreciprocal ferrite phase shifter unless the slab is moved away from the wall.[2]

The Ferrite Dielectric Mode

Another mode, not associated with a normal waveguide mode, also propagates at a large value of β shown by curve C. The electric field for this mode shown in Fig. 2. There is no propagation in the reverse direction in this mode until the waveguide is almost filled with ferrite. At this point forward propagation cut off and propagation in the reverse direction is established as shown by mode D.

Fig. 2—Propagation constant of lossless ferrite modes in standard 9,000 mc rectangular waveguide as a function of ferrite slab thickness. The single slab against the wall has a saturation magnetization of 3,000 gauss. The internal dc field indicated is 1,000 oersteds.

The existence of these two sets of mode patterns represented by A-B and C-D is significant for the application of these modes to practical devices. In particular, referring to Fig. 2, it can be seen that in large range of ferrite thickness, propagation is exclusively in mode C which is unidirectional. The theoretical results indicate the possibility of one-way propagation. It necessary, however, to use a sufficient thickness of material to cut off the reverse propagation of mode D. Furthermore, as one might expect, this cutoff depends strongly on waveguide width, frequency and applied field.

On the assumption that this one-way transmission line is a two-port device, thermodynamic arguments can be advanced which indicate that such a device cannot exist for a lossless idealized system. Further investigation will be made to resolve this apparent paradox.

---

[2] B. Lax, K. J. Button, and L. M. Roth, *Jour. Appl. Phys.* vol. 25, p. 1413; 1954.

The Ferrite-Guided Mode

In attempting to analyze mode C in the limit of small thickness, the analysis indicated that it could not be associated with any of the normal modes of the guide. The electric field configuration of this mode strongly suggests that it is analogous to dielectric propagation which has been discussed by Schelkunoff.[3] Consequently we consider this possibility directly; that of a semi-infinite ferrite slab backed by a conducting plate as shown in Fig. 3. Here one can assume that the electric field within the ferrite follows an hyperbolic sine and that it is decaying exponentially in the region outside.
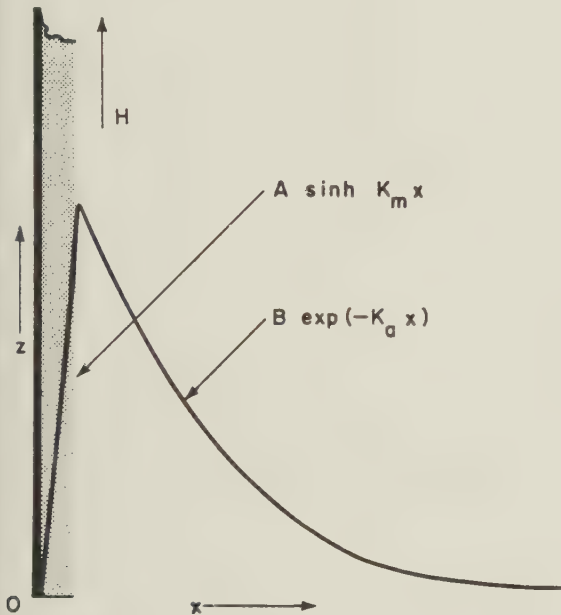
Fig. 3—A semi-infinite ferrite slab backed by a conducting plate. The electric field for the unidirectional guided mode is shown. The propagation constant as a function of slab thickness is shown as a dashed line on Fig. 2.

This field is very similar to that of mode C. The substitution of these fields into Maxwell's equations, following the procedure outlined above, results in the following transcendental equation which has been solved for the slab width:

$$\delta = \frac{1}{K_m} \tanh^{-1}\left(\frac{jK_m\mu_o}{\frac{\beta\mu_o}{\theta} - k_a\mu_{eff}}\right) \qquad (6)$$

The positive values of β for forward propagation yield a mode guided by the ferrite as assumed in Fig. 3. The dashed line of Fig. 2 shows the propagation characteristic which coincides with the line for mode C for a waveguide less than half filled. Apparently these are the "ferrite dielectric" modes. However, negative values of β for reverse propagation do not yield a guided mode. Therefore the unidirectional behavior of this mode in a waveguide is not surprising.

[3] S. A. Schelkunoff, "Electromagnetic Waves," p. 428, D. Van Nostrand Co.; 1943.

THE TWIN-SLAB STRUCTURE

The Reciprocal Mode

Another and closely related problem is that of twin slabs, one against each side wall of the guide and each magnetized in the same direction. Although the equations are slightly more complicated by the addition of two boundary conditions, the method of analysis is exactly the same as before and leads to the solution:

$$L = 2\delta + \frac{1}{k_a} \tan^{-1}\left(\frac{p}{q}\right), \qquad (7)$$

where

$$p = \frac{2K_m}{k_a} \cosh K_m\delta \sinh K_m\delta$$

and

$$q = \left[1 - \left(\frac{\beta\mu_o}{\mu_{eff}\theta k_a}\right)^2\right]\sinh^2 K_m\delta - \left(\frac{\mu_o K_m}{\mu_{eff} k_a}\right)^2 \cosh^2 K_m\delta.$$

The absence of the linear term in β indicates that propagation associated with this structure is reciprocal. Suitable sets of values of β and δ are chosen until (7) yields a guide width L corresponding to standard 9,000 mc waveguide. These results are shown in Fig. 4 where
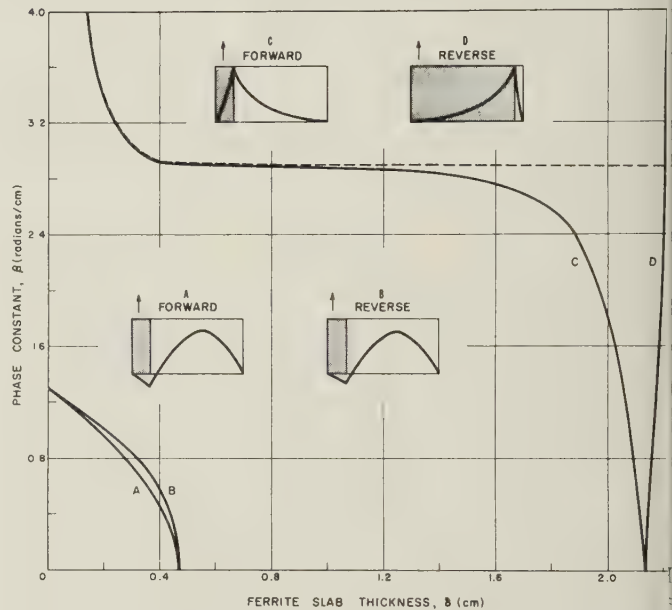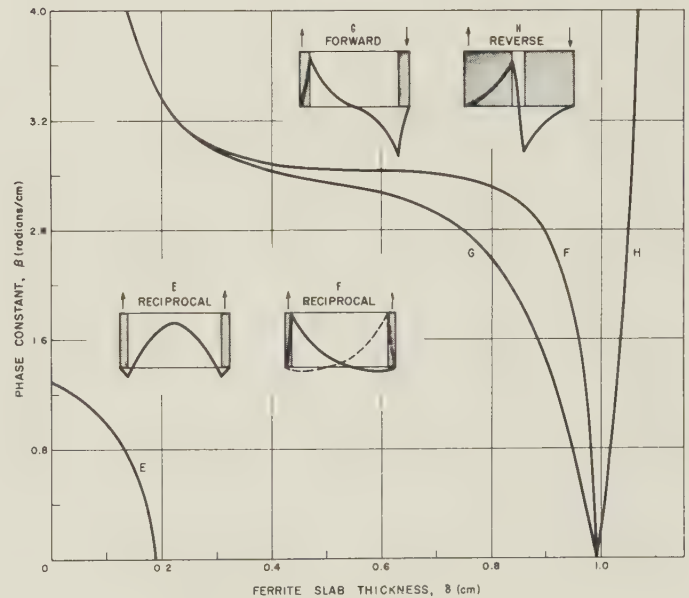

Fig. 4—Propagation constant of the twin-slab lossless ferrite modes in standard 9,000 mc rectangular waveguide as a function of the thickness of each slab.

mode E, associated with the dominant waveguide mode, is again cut off through the use of relatively thin slabs. Since mode E is reciprocal and is almost symmetrically distorted, there is no obvious device application here. It is the ferrite dielectric mode that is again of interest for practical purposes. The propagation characteristic, shown by curve F, is reciprocal but the direction of propagation determines which slab shall "guide" the wave. For forward propagation, the electric field has a very sharp peak at the face of the left slab. In the other direction the peak is at the face of the other slab as

shown by the dashed line. This explains the principle of operation of the resistance sheet isolator discussed by Fox, Miller and Weiss.[4] This device depends for its isolation on the presence of a resistance card against the face of one of the slabs so that a large loss is experienced in the direction in which the card finds itself in high E-field. In the opposite direction, the attenuation should be much less since the electric intensity is very small.

The Nonreciprocal Mode

By reversing the magnetic field at one of the ferrites of the twin slab structure, a third set of modes can be achieved. Here again the ferrite dielectric mode is nonreciprocal and unidirectional as shown by curves G and H in Fig. 4. Mode G propagates only in the forward direction and is cut off when the guide is nearly completely filled. Reverse propagation in mode H is then established. As before, the theory indicates that a proper choice of slab thickness may result in a one-way transmission line. The electric fields for modes G and H are antisymmetric with a node at the center of the guide. Eq. (5) for nonreciprocal modes was used to obtain these results where L was taken to be half the actual waveguide width to give the electric node at the center. The lowest nonreciprocal TE waveguide modes for oppositely magnetized twin slabs are not shown here because they would be expected to follow curve E fairly closely.

Magnetic Field Configurations

The rf magnetic field patterns for these modes are also easily obtainable from this analysis and are helpful in the discussion of the behavior of other devices. The longitudinal component, h_y, and the transverse component, h_x, are shown separately in Fig. 5 for the forward propagation of the reciprocal mode F. Note that the transverse component of $\vec{h}$ crosses the axis and becomes negative at about $3L/4$. The longitudinal component of $\vec{h}$ however, never crosses the axis. For thicker slabs, h_y does not achieve the low values that it does in the case shown here. It is the longitudinal component which determines the efficiency of operation of the slotted-wall field-displacement circulators discussed by Fox, Miller and Weiss.[5] These devices operate on the same principle used in a simple slotted line where a longitudinal nonradiating slot is cut in the broad face of the guide along a line where the longitudinal h-field is zero. In the case of the ferrite loaded guide, even though the h_y field is not zero, it is so small (for a thin slab) on the right hand side that very little radiation would be observed through a longitudinal slot cut in this region. But for reverse propagation of the reciprocal mode F, the peak of the h-field is found on the right-hand side so that radiation would take place through a slot on this side. For optimum performance it appears from Fig. 5 that the maximum directivity or isolation in the devices described would be obtained if the slot were placed near the face of the slab. Furthermore, since
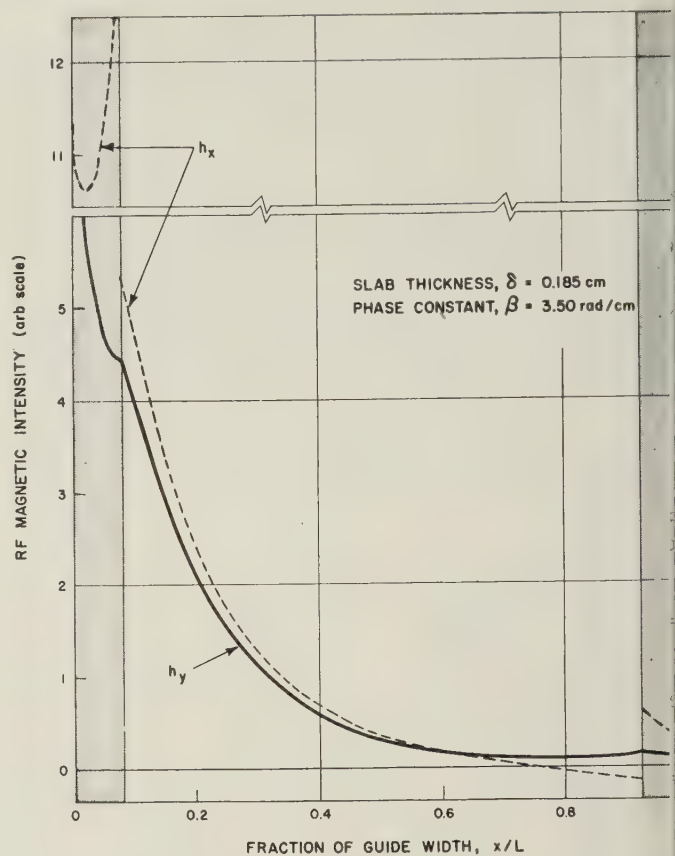
Fig. 5—RF magnetic field components of the twin slab mode for forward propagation. This ferrite dielectric mode is reciprocal but the peak intensities will appear on the right for reverse propagation.

a large ratio of maximum h_y to minimum h_y is essential to the successful operation of this circulator, a thin slab is required. However, sufficiently thick slabs must be used to cut off the waveguide mode E.

The rf magnetic field patterns for the single slab ferrite modes are almost identical to those of the twin slab modes except that neither component of $\vec{h}$ goes negative anywhere.

THE NONRECIPROCAL PHASE SHIFTER

Magnetic Field below Ferromagnetic Resonance

If one is not restricted to operation of the device with slab against the wall but will permit the slab to be moved away from the wall, the solutions appear to be quite different. The analysis in this case is considerably more complicated by the addition of the parameter a, the distance of the slab face from the nearer waveguide wall. A drawing of the slab in the guide is shown in Fig. 6. A sine or hyperbolic sine function is suitable to describe the rf electric fields in regions 1 and 2 but a linear combination of a sine and cosine (or hyperbolic functions) is required in the ferrite region 3. The method of analysis and numerical solution, however, is similar and leads to a transcendental equation, which in this case, has been solved for the position of the slab.[6]

[4] A. G. Fox, S. E. Miller, and M. T. Weiss, *Bell Sys. Tech. Jour.*, vol. 34, pp. 5–104; 1955.

[5] *Ibid.*

[6] Lax, Button, and Roth, *op cit.*
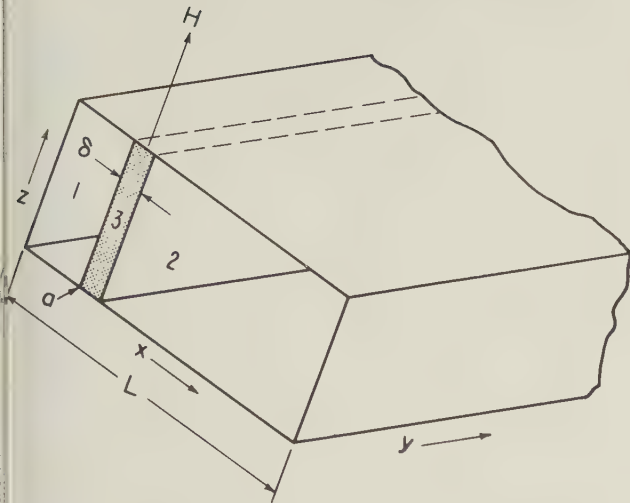
Fig. 6—A single ferrite slab moved a distance a from the wall of rectangular waveguide of internal width L.

$$= \frac{L - \delta}{2} - \frac{1}{2k_a} \cos^{-1}\left[\frac{-pr \pm q(q^2 + p^2 - r^2)^{1/2}}{p^2 + q^2}\right], \quad (8)$$

where $p = \frac{1}{2}\left(\frac{k_a^2 \mu_{eff}^2}{\mu_o^2} + \frac{\beta^2}{\theta^2} - k_m^2\right)$

$$= j\frac{\beta k_a \mu_{eff}}{\theta \mu_o}$$

$$= \frac{1}{2}\left(\frac{k_a^2 \mu_{eff}^2}{\mu_o^2} - \frac{\beta^2}{\theta^2} + k_m^2\right) \cos k_a(L - \delta)$$

$$+ k_m \cot(k_m\delta)\frac{k_a\mu_{eff}}{\mu_o}\sin k_a(L - \delta).$$
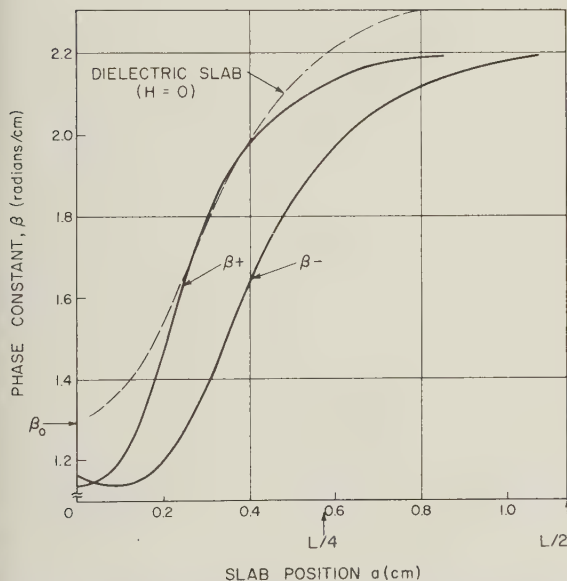
Fig. 7—Phase constant vs position of slab for a slab thickness 0.1 cm. The dotted line shows the variation of β for a nonmagnetic dielectric slab having the same dimensions and dielectric constant the ferrite slab.

his equation has been solved numerically for a slab the ickness of one millimeter and the result is shown in g. 7 for the lowest TE mode. Here the points at $a = 0$ orrespond to the points at $\delta = 0.1$ cm on curves A and of Fig. 2. If the slab thickness were increased, the intercepts of Fig. 7 would move toward $\beta = 0$ as scribed by curves A and B of Fig. 2. But if the slab
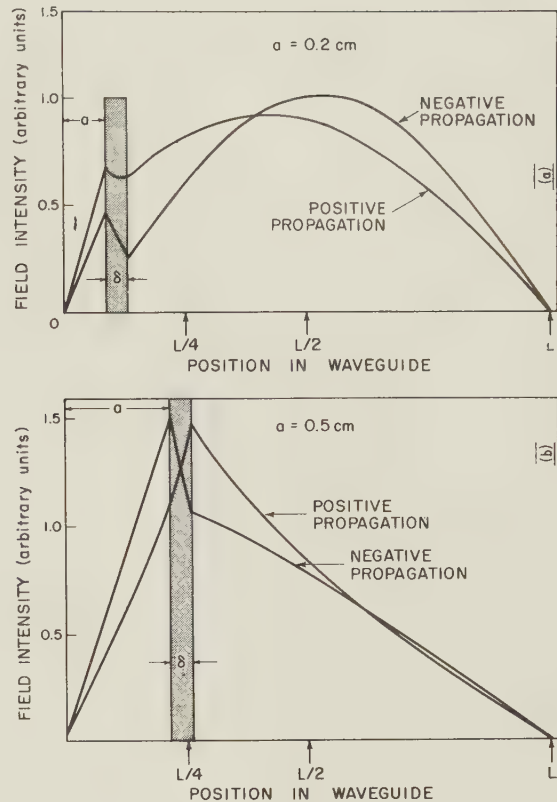
Fig. 8—Relative electric field intensity as a 0.1-cm ferrite slab is moved away from the wall. The internal dc field is 1,000 oersteds and the frequency is 9,000 mc. The field pattern for a nonmagnetic dielectric slab would be similar in the empty region to the curve for positive propagation in the two cases shown.

is moved away from the wall the phase constant increases sharply. Access to the cutoff condition of the waveguide mode is thus denied except for very small values of a, or possibly, for very thick slabs.

This can be understood, perhaps, by noticing that as the slab is moved away from the wall, a continuous field displacement takes place, as shown by the electric fields of Fig. 8. Here the waveguide width appears to be effectively enlarged rather than being reduced. This enlargement effect finally reaches a point at $a \simeq L/4$ when the lowest TE mode gradually turns into a ferrite dielectric-type mode. The latter is characterized by an hyperbolic function for the transverse component of the rf electric field in the empty region.

The work carried out on this problem has been helpful in understanding the nonreciprocal phase shifter.[7] The differential phase shift, $\beta_+ - \beta_-$, can be obtained directly from curves like those of Fig. 7. It is well known now that the proper position of the slab for maximum differential phase shift is closer to $L/8$ than to $L/4$ for practical values of δ as shown in Fig. 9. These results are for several thicknesses of material at an internal dc field of 200 oersteds. As the slab thickness is increased, the optimum position of the center of the slab moves to the left until the slab face is nearly against the wall. Further discussion of this nonreciprocal phase shifter and also the lowest TE waveguide mode for twin slabs magnetized oppositely may be found in reference 2.

[7] Ibid.
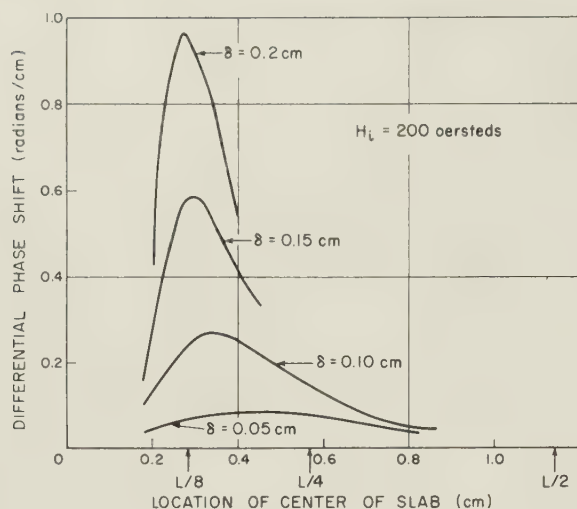
Fig. 9—Differential phase shift vs location of center of ferrite slab for several values of slab thickness.
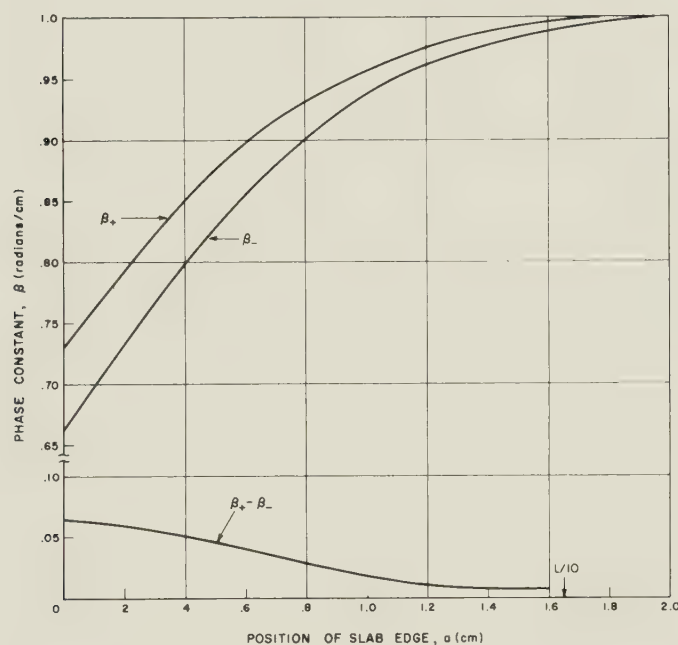
Fig. 10— The phase constants for forward and reverse propagation (upper figure) and the differential phase shift (lower figure) as a function of slab position. This solution is for a slab 1.35-cm wide in a standard 1,300 mc rectangular waveguide having an internal width of 16.5 cm. The ferrite has a saturation magnetization of 3,000 gauss and is operated above ferromagnetic resonance at an internal dc field of 1,000 oersteds.

Magnetic Field Above Ferromagnetic Resonance

The region above ferromagnetic resonance may be of interest for propagation in ferrite-loaded waveguides at low microwave frequencies. The assumption of a low loss ferrite is a bit more difficult to justify in this frequency region. However, if one assumes that it is possible to obtain polycrystalline ferrites which have line widths of 50 oersteds (equivalent to single crystal material) then these calculations which are based only on the dispersive properties should be valid.[8] An evaluation of the differential phase shift for the single slab case has been carried out at a point several hundred oersteds above resonance for a frequency of 1,300 mc. The result
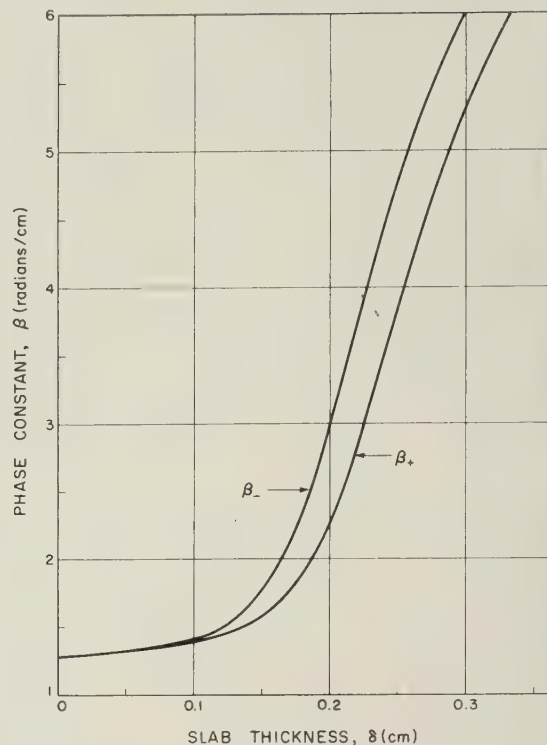
Fig. 11— Phase constant vs ferrite slab thickness for forward and reverse propagation in the region above resonance. This solution is for a slab against the wall in standard 9,000 mc waveguide. The internal dc field is 4,000 oersteds.

has been obtained as a function of slab position using (8). The upper curves of Fig. 10 show the phase constants for a slab thickness of 1.35 cm (about 8 per cent of the guide width).

Since the differential phase shift, shown in the lower curve, indicates that the optimum position for this thick slab is against the wall, it is worth while to return to the simple equation which we have for this case, (5), to discover the nature of the solutions obtained above ferromagnetic resonance. The curves shown in Fig. 11 have been worked out for a single slab against the wall using an internal dc field of 4,000 oersteds at X-band. Here the "cutoff" of the waveguide modes is not accomplished at all. A continuous field displacement takes place as β increases with δ until each mode gradually becomes a ferrite dielectric mode. The electric fields are shown in Fig. 12 for forward and reverse propagation at the transition point where forward propagation is still in the lowest TE mode but the reverse propagation is guided by the ferrite.

Nonreciprocal Attenuation

Each of the problems discussed thus far have been restricted to regions of negligibly small attenuation. It is possible, however, to extend the calculations to include losses, in which case, the propagation constant becomes complex, i.e., $\Gamma = \alpha + j\beta$. The particular loss mechanism which has been investigated is that associated with ferromagnetic resonance and the complex permeability tensor is represented in terms of the Bloch-Blombergen relaxation time.[9] If these quantities are substituted in

---

[8] B. Lax, "A figure of merit for microwave ferrites at low and high frequencies," *J. Appl. Phys.*, vol. 26, p. 919; 1955.

[9] Lax, *op cit.*
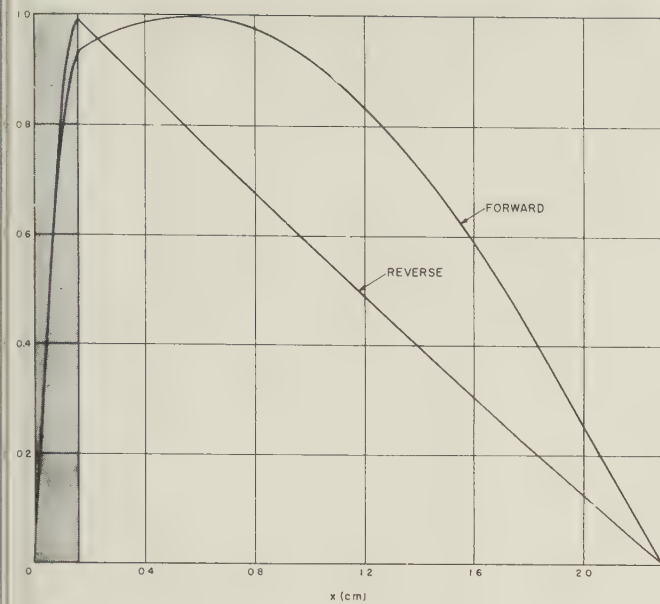
Fig. 12—Electric field intensity in a standard 9,000 mc wave-guide with a single ferrite slab against one wall. The internal dc field 4,000 oersteds. The slab thickness of 0.16 cm has been chosen uch that forward propagation is in the lowest TE mode while reverse propagation is guided by the ferrite.

Fig. 13—Attenuation and phase constants for each direction of propagation as a function of the location of the slab edge. The slab thickness is one millimeter and the properties of Ferramic R-1 at 9,375 mc were used. The internal dc field is 2,100 oersteds.

8), two simultaneous transcendental equations can be btained. Choosing suitable values of α, β and δ, the orresponding position of the slab for the lowest TE mode can be found. Typical curves of the attenuation nd phase constant as a function of slab position are hown in Fig. 13. These calculations were carried out at ,375 mc for a slab one millimeter thick using the X-band roperties of Ferramic R-1. The resonance line half-width at half-maximum is 250 oersteds. The point A marks the location of the edge of the slab for maximum ifferential attenuation, and B is the point of maximum ifferential phase shift.

The internal dc field used here is 2,100 oersteds, which s 400 oersteds below resonance. Fig. 14 shows the plot f the maximum forward attenuation vs internal dc ield. The curve was not carried above the half-maximum oint because of mathematical complication.

CONCLUSION

The analysis has been carried out for a few structures nd a restricted choice of parameters to illustrate some f the properties and applications of the solutions. These do not exhaust the possibilities of this theory. The properties described here are certainly sensitive to hanges in the magnitude of the internal dc field and robably also to the ferrite saturation magnetization. Frequency dependence within several bands of fre-quencies must be investigated. The unidirectional modes nd cutoff conditions suggest possibilities for circulators, ilters and radiation and switching devices. Changes in arameters may lead to optimum design conditions for uch applications.

The structure chosen for the demonstration of attenua-ion calculations is a fairly complicated one. It would be rofitable to repeat such calculations for one of the
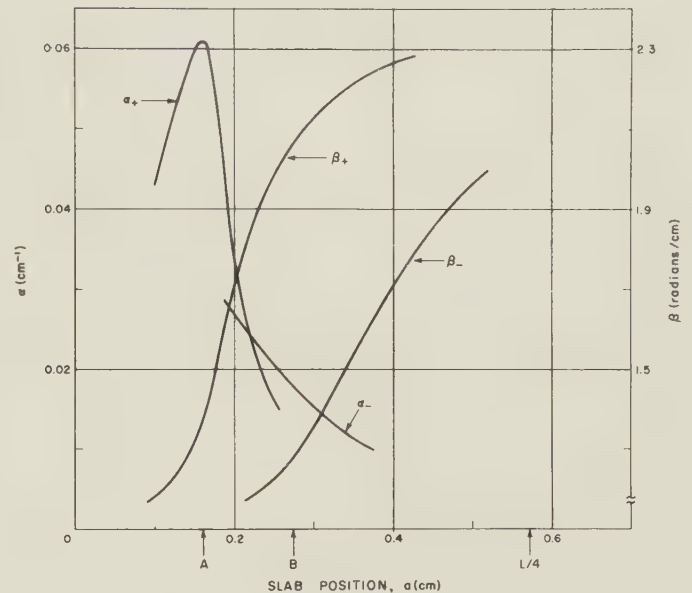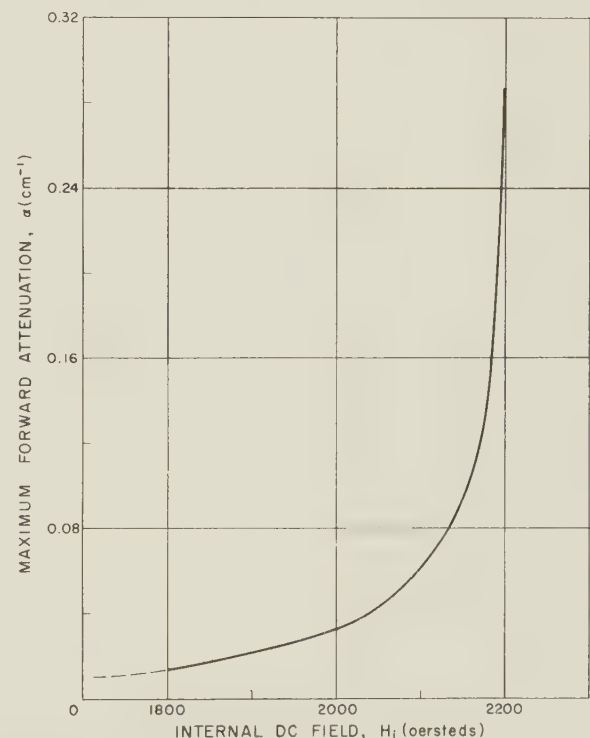
Fig. 14—Maximum forward attenuation versus internal dc field in a region below ferromagnetic resonance.

simpler structures in the investigation of the ferrite dielectric modes. Dimensional effects observed in wave-guide and cavities can be studied in this way and some are probably associated with these new modes.

Panel Discussion on Boundary Value Problems of Diffraction and Scattering Theory—(I)

Chairman: G. Sinclair

Panel Members

N. Marcuvitz	W. Franz	R. K. Ritt
C. Müller	R. Timman	W. Braunbek
I. Imai	C. Schensted	H. C. van de Hulst

G. Sinclair, Univ. of Toronto, Toronto, Can.: When he was organizing this Symposium, Dr. Siegel attempted to select papers which were on a high scientific level. However, I am glad that he did not exclude engineers, as many engineers have a very legitimate interest in the problems being discussed since, for example, the solution to a diffraction problem is potentially useful as the solution to an antenna problem. In constructing antennas, it is usually necessary to mount them near a structure such as a building, ship, or aircraft. The calculation of the receiving pattern of the antenna in the presence of the structure can often be reduced to a determination of the diffracted or scattered field at the antenna when a plane wave is incident on the structure. Hence I suggest that we do not lose sight of the fact that there may be a practical application, no matter how mathematical the paper.

W. S. Ament, Naval Res. Lab., Washington, D. C.: Bremmer and others have shown that normal mode solutions in microwave problems and ray solutions are connected through Poisson's summation formula. It is my opinion that the normal Fourier-Bessel expansion, for cylindrical diffraction problems at least, and the creeping wave solution are similarly connected. The kinds of integrals one gets look very much like Ramanujan's integrals on the order of Bessel functions, and Poisson's summation procedure can be applied if it is possible to take a Fourier transform on the order of the Bessel functions. What is the connections between this and Watson's transformation?

W. Franz: Münster Univ., Ger.: The Watson transformation is fairly simple to use, and if we use it, we can make several of the series more convergent.

C. Müller, Univ. of Bonn, Ger.: In order to justify the application of Poisson's summation formula you have to check quite complicated conditions, whereas the application of Watson's transformation is straightforward. It may turn out in the end, that what you get by using Poisson's summation formula is what Dr. Franz gets by his contour integrals.

F. J. Zucker, AF Cambridge Res. Center, Cambridge, Mass.: Dr. Marcuvitz has given an example in which there is a continuous spectrum of leaky waves. The structures we deal with in engineering quite frequently do not have a continuous spectrum, but a discrete one, where you have infinitely many discrete leaky modes. In ordinary waveguides, or slow open waveguides we can manage to excite one leaky mode predominately. Since you don't have a wedge or cone, in what domain are these leaky waves strictly applicable?

N. Marcuvitz, Polytechnic Inst. of Brooklyn, Bklyn, N. Y.: Leaky waves can only arise if there exists a continuous spectrum of modes, and, in fact, leaky waves are alternative means of expressing radiation from the system.

Zucker: I realized that the leaky waves had to equivalently represent a continuous spectrum, but I didn't realize that they had to form one themselves. If you apply the transverse wave equation with a transverse impedance condition, and allow a complex propagation constant, then you get, in certain cases, a discrete of solutions. This is not a continuous set.

Marcuvitz: I didn't mean to imply that the leaky wave solution, being the poles of the reflection coefficient or the zeros of the impedance of the system, always form a discrete set without exception. They are always discrete in number but they may be infinitely discrete. In the case treated, namely the case of a structure with no dissipation, in the form of a channel in a non-dissipative dielectric medium, these complex zeros are infinite in number but most of them correspond to beyond-cutoff waves, that is waves that leaked but nevertheless were beyond cutoff.

J. B. Keller, New York Univ., New York, N. Y.: The problem of the corner of a screen has also been treated by Kraus of New York University, by means of separation of variables in sphero-conal coordinates. There then results the problem of the eigenvalues of a pair of simultaneous Lamé equations, which is treated to a certain extent but not completely. One finds that there are three explicitly known eigenvalues, namely, 0, $\frac{1}{2}$, and 1, corresponding to screen angles of 0 (a needle), π (a half plane), and 2π (complete plane). From this, Levy of California conjectured that the eigenvalue is a linear function of the screen angle, and, in order to verify the conjecture, he computed the derivative of the eigenvalue with respect to the angle at the angle π. He found the value of the derivative was indeed one, in accordance with the conjecture. What is the extent of the disagreement between Levy's conjecture and the results obtained by Prof Braunbek? I might also mention that the results obtained by Kraus, which are not absolutely precise as yet, seem to be in disagreement with the conjecture.

W. Braunbek, Univ. of Tubingen, Ger.: My method is only an approximate one, and cannot be used when the angle is very small, or very near π, so the case of a needle cannot be calculated with my method. My method can only be used for angles near the value $\pi/2$. I cannot tell at this moment whether my results agree or disagree with the conjecture.

H. Motz, Oxford Univ., Eng.: I want to ask Prof. Marcuvitz whether this particular representation of solutions with reflection coefficients is confined to the two-dimensional case which he treated. For example, with cylindrical symmetry, the solutions could, I imagine, not be expressed in this form. What would be the relation to the poles of the reflection coefficient?

Marcuvitz: I believe that in three-dimensional problems, and particularly in the case of cylindrical symmetry which you mention, exactly the same situation prevails. There is an exact analogy between that problem and the two-dimensional problem which I treated, the only difference being that the $e^{in\gamma}$ functions are replaced by combinations of Bessel and Neumann functions. In fact the field produced by a point source in a cylindrical region of the type I considered has been worked out in an Indian journal. The problem treated may have been a scalar one, but in any case the vector problem can be solved.

V. Twersky, Sylvania Elec. Prods. Inc., Mountain View, Calif.: Dr. Müller indicated he could compute a certain function, called the efficiency, for a distribution of dipoles, and this was essentially a normalized maximum intensity output. It was not clear to me that you could actually compute the maximum value of this, without knowing explicitly the amplitudes of the dipoles in the distribution, that is, without knowing the actions and interactions of one dipole with another.

Müller: Perhaps I had better answer this by giving an example of a scalar case, in which the matrix has roots λ_1 and λ_2. Without knowing the problem, we can determine the efficiencies which are the values of λ_1 and λ_2 directly, just by knowing the geometry of the arrangement. The same is true in the vector case, except that the matrix is more complicated, but again is determined by the geometry of the set-up.

Keller: In the case of creeping waves on a perfectly conducting cylinder, one of the most important quantities is the decay factor or the exponent that characterizes the

mode. Has Dr. Franz computed these mode constants for waves on a dielectric cylinder?

Franz: We are now making these computations.

H. C. Van de Hulst, Leiden Univ., Netherlands: Can anyone inform me of any numerical work which has been done for diffraction by a parabolic cylinder?

Keller: S. O. Rice published some work in the *Bell System Technical Journal* some time ago. He used the exact solution for a perfectly conducting parabolic cylinder, in terms of parabolic cylinder functions and took a Watson transformation. His results can be interpreted in terms of creeping waves, as I will show in my paper later in this Symposium.

C. L. Pekeris, Weizmann Inst., Rehovoth, Israel: I would like to ask a question of Dr. Müller. In the case of a continuous distribution of m, if you have a closed medium, simply connected, with a continuous m, and you observe the electromagnetic field on a surface outside it, can you from that, determine m?

Müller: No. To determine the coefficient of diffraction from the diffraction itself, is, as far as I can see, far from being solvable at present.

Pekeris: When you say "no," do you mean the problem is not unique?

Müller: It may be. It can be shown there are changes such that they can be inside the region and not radiate at all.

Twersky: Would Dr. Franz explain the physical characteristics of creeping waves?

Franz: Creeping waves are analogous to water waves incident on a wooden cylinder. I might also mention that Hartnagel has done some experiments on diffraction by a razor blade using optical wavelengths and has obtained good agreement with the creeping wave concept.

Van de Hulst: The creeping wave concept can be applied to any curved surface, and besides it is not necessary to think of the injection as due to a plane wave.

Panel Discussion on Boundary Value Problems of Diffraction and Scattering Theory (II)

Chairman: S. Silver
University of California
Los Angeles, Calif.

Panel Members

H. Bremmer*	A. E. Heins*	M. Kline*
P. C. Clemmow*	D. S. Jones*	E. K. Ritter*
C. L. Dolph*	J. Keller*	K. M. Siegel*

W. V. Tilston (Sinclair Radio Labs., Toronto, Can.)

The discussion period was a lively one in which members of the audience as well as those of the panel participated. The following is reconstructed from a tape recording of the proceedings (with one exception as noted) and at least the general sense of the remarks is given here. In view of the fact that the speakers supplemented their comments by blackboard material of which there is no record and, as time does not permit submitting the transcriptions to the respective contributors for checking, it seems inadvisable to attempt to report the discussions in the form of direct quotations. The names of contributors to the discussion who were not members of the panel are listed under each paper.

1. **Kline—"Electromagnetic Research at the Institute of Mathematical Sciences of New York University"**

Contributor: S. Karp, New York University. Prof. Kline, in presenting his paper, chose some six topics as being representative of the character and scope of the program of the Electromagnetic Research Project. In response to inquiries during the discussion period concerning the distribution of these topics among the group, Kline listed the individual members of the project particularly responsible for the work. Since his paper is a comprehensive survey of the entire project and is well annotated by appropriate references there seems to be no point in giving references for the particular subjects which he spoke about as illustrative material.

Among the topics which Kline reviewed was the work by Karp on the relationship between the separation of variables technique and the Wiener-Hopf integral-equation technique for solving certain classes of boundary value problems. In this connection Heins inquired as to the sense in which the separation of variables technique was construed for treating the radiation (or scattering) from a pair of semi-infinite parallel conducting planes. Karp pointed out that, in general, the class of problems which he treated is that for which separation of variables in cartesian coordinates can be used by division of the space into suitable regions. Thus, in the case of the parallel planes three fundamental regions are considered, one the semi-infinite free region, a second the space between the plates, and the third the two partially infinite regions in the semi-infinite domain containing the plates which are exterior to the parallel plate region. The solutions are set up in the three regions in terms of basic functions which arise by separation of variables in cartesian coordinates. The process of connecting the three regions then leads to a function-theoretic problem of the same basic nature as that of solving the Wiener-Hopf integral equation for the system.

2. **Bremmer—"Asymptotic Developments and Scattering Theory in Terms of a Vector Combining the Electric and Magnetic Fields"**

Keller inquired as to the special advantages of using the Bateman-Silberstein complex vector inasmuch as it is necessary to resolve the vector into its component electric and magnetic fields when treating the boundary value problem. Bremmer pointed out that as in the Luneberg-Kline work he was treating the general theory of asymptotic solutions to Maxwell's equations without regard to a special type of boundary value problem such as is encountered in scattering by a perfectly conducting obstacle. The Bateman-Silberstein representation leads to a simultaneous development of all higher-order equations for both the electric and magnetic vector and thus makes possible the discussion of the electromagnetic field as an entity. If one wishes to relate the development to a boundary value problem, the appropriate problem he should consider is that of determining the asymptotic form of a field corresponding to assigned values of E and H over a surface, the objective being to determine the propagation of the field in the sense of transport along a family of rays. When dealing with a scattering problem it is indeed necessary to resolve the Bateman-Silberstein vector into its component fields E and H and, as Kline pointed out, it is further necessary to resolve each field vector into spatial components in order to match the boundary conditions.

3. **Siegel, Schultz, Gere and Sleator—"The Theoretical and Numerical Determination of the Radar Cross Section of a Prolate Spheroid"**

Contributors: H. C. van de Hulst, Leiden University, W. Franz, Munster University, L. I. Schiff, Stanford University, and F. J. Tischer, Redstone Arsenal. Siegel's point that the particular oscillatory character of the variation of the radar cross section of the prolate spheroid with wavelength, in which the second maximum is considerably greater than the first, makes the applicability of creeping wave theory questionable stimulated a great deal of discussion on the prolate spheroid problem and creeping waves in general.

Tischer attempted to show that the creeping wave part of the solution to diffraction by an obstacle is an explicit result of just Maxwell's equations. He showed that the geometrical shadow is inconsistent with the field equations, and that there is a field in the shadow region which is an analytic extension of the field in the illuminated region of space. However, his analysis cannot delineate explicitly the character of the field in the shadow region; the creeping wave phenomenon cannot be extracted from just the field equations alone.

The question of treating the scattering by the spheroid as interference between an optical field and a creeping wave field was taken up by van de Hulst. He first called attention to the fact that for the case treated by Siegel and his co-workers in the region of the scattering curve under consideration, the dimensions of the spheroid are such that the radius of curvature at the nose is $kR = 0.014$. The creeping wave theory is an asymptotic representation of the field which is obtained when the dimensions are large compared with the wavelength and, therefore, it should not be expected to yield significant results when the dimensions are small compared with wavelength. (It should be noted here that Siegel's consideration of applicability of creeping wave theory was based on the point of view that although its formulation is obtained out of an asymptotic solution, it represents a physical process and, therefore, it is desirable to investigate its applicability to problems such as treated in his paper.)

Prof. van de Hulst proposed that the scattering from the spheroid should be considered from an entirely different point of view. Taking into consideration the fact that the transverse dimension of the spheroid is small compared with the wavelength, he suggests that the structure be treated sectionally by static field approximations such as used by Rayleigh in the treatment of scattering objects entirely small compared with the wavelength. A phase distribution along the major axis of the spheroid is to be introduced to correspond with the fact that in this direction the dimension is comparable with the wavelength. For a fixed axial ratio, as the ratio of major axis to wavelength increases, the transverse dimension increases proportionally and more coherent radiation returns in the direction toward the oncoming wave. It is thus possible for the second maximum to exceed the first. Lastly, he pointed out that a more critical evaluation of the phenomena can be made by studying the complex amplitudes rather than just absolute values and he should like to see the complex scattered amplitude as function of the dimensions.

In a written communication Schiff develops a similar idea. The following is the complete text of Schiff's communication:

"The purpose of these remarks is to propose a simple physical explanation for the striking difference between the back scattering of electromagnetic waves from a sphere and from a prolate spheroid viewed end on. When the cross section is plotted against $2\pi a/\lambda$, where a is the radius of the sphere and the semi-major axis of the spheroid, the first few successive maxima and minima come in about the same places. When expressed in units of πa^2 for the sphere and $\pi b^4/a^2$ for the spheroid, where b is the semi-minor axis (these being the asymptotic values of the cross sections), the first maximum and the first minimum have roughly the same values. Succeeding maxima decrease and succeeding minima increase in the case of the sphere. However, the first few succeeding maxima increase in the case of the spheroid, and the corresponding minima either remain about the same or decrease. It seems likely that the comparison of the minima is not very significant, since they are all close to zero anyhow (this does not show well on a logrithmic plot). The comparison of the maxima, on the other hand, calls for a physical explanation.

It is suggested that a fundamental difference between the sphere and spheroid cases is the following. In the case of the sphere, the dimensions along and transverse to the direction of wave propagation are the same. Thus retardation effects, which arise from comparable magnitudes of object dimensions and wavelength, appear at the same value of $2\pi a/\lambda$ for both transverse and longitudinal directions. In the case of the spheroid, on the other hand, retardation becomes significant for the longitudinal direction at a value of $2\pi a/\lambda$ somewhat less than unity, whereas at this point $2\pi b/\lambda$ is very much smaller than unity. Thus one must consider interference effects along the length of the spheroid long before one need consider them for the transverse dimensions. So long as the transverse dimensions are small compared to a wavelength, the scattering effects of the object can be represented by an oscillating electric or magnetic dipole; in this regime the cross section rises in proportion to λ^{-4} as λ decreases or $2\pi a/\lambda$ increases, and is referred to as Rayleigh scattering. When the wavelength becomes comparable with the longitudinal dimension of the object, interference maxima and minima appear because of the phase difference between the currents induced in portions of the object at different distances from the source. In the sphere case, the maxima and minima and the departure from Rayleigh scattering appear at the same point, since the longitudinal and transverse dimensions are the same; thus as soon as the scattering shows interference oscillations, it tends to settle down to its asymptotic value. In the 10:1 spheroid case, the oscillations appear first as λ decreases, but the Rayleigh rise is expected to continue until $2\pi b/\lambda$ becomes comparable with unity, after which the cross section should settle down to its asymptotic value. The spheroid calculations have not yet been calculated out far enough to see whether or not this behavior is found.

The validity of the foregoing picture could be tested by calculating the back scattering from a spheroid that has on its surface a current distribution defined in the following way. Imagine the spheroid to be immersed in

a static electric field perpendicular to the major axis, and along the direction of polarization of the incident radiation. This will give rise to a surface charge distribution. Then allow the field to oscillate sinusoidally in time with very low frequency; this will give rise to a surface current density that also oscillates in time with the same frequency, and that has a dipole character so far as the short (transverse) dimensions of the spheroid are concerned. Finally, assume that this current distribution is modulated by a phase factor that is a function only of the distance parallel to the major axis, and that corresponds to the actual wavelength. Such a calculation would not be very laborious, although not trivially easy, either.

Since the calculation described in the last paragraph has not yet been done, the result of a much simpler calculation will be mentioned here. The spheroid is divided into infinitesimal sections by planes perpendicular to the major axis. The static surface charge density is computed for each of these sections as though it were part of an infinitely long cylinder of the local diameter immersed in a transverse electric field. Then as the field changes slowly, the surface current density is computed as though the charge flows only circumferentially, not along the length of the spheroid (which would actually be the case since the diameter changes from point to point). The back scattering is then computed from this current density, modulated by a phase factor along the spheroid. The result of this very simple calculation reproduces the numerically computed back-scattering cross section, except for: (1) the Rayleigh rise and the first maximum are one-quarter of the correct values; (2) successive maxima are equal instead of increasing; (3) successive minima are zero. Maxima and minima appear in about the correct places. It is expected that the slight shadowing of the half of the spheroid away from the source will prevent the minima from being zero, which would improve agreement with the exact calculation. Further, the calculation described in the last paragraph must certainly give the correct Rayleigh rise; since this is well known to be given correctly in terms of the static dipole moment; thus the factor of four would come out correctly. Such an improved calculation should therefore take care of points (1) and (3) above; it does not seem possible to say in advance whether or not it will take care of (2) as well."

Franz turned the discussion once more to the creeping wave model emphasizing the point that the creeping wave theory is an asymptotic result obtained when dimensions are large compared with the wavelength. He was sympathetic with the view expressed by Siegel that the creeping wave solution is possibly more than just an asymptotic representation and has a basic physical significance, but noted that Siegel had mistakenly considered the creeping wave contribution to be monotonically dependent on the dimensions. In the case of a sphere the creeping wave contribution behaves like

$$(ka)^{4/3} e^{-2.2(ka)^{1/3}}$$

which is a function that exhibits a maximum. The damping factor is somewhat different for the case of a sphere from that for the case of a cylinder and should again be different for the spheroid. Franz felt that the damping factor should be closer to 3 than to 2, because the nose in the case under consideration is more like a sharp edge than a smoothly varying cylindrical surface. By combining the several factors he was able to show that the creeping wave analysis gives the position of the observed minima with good accuracy (or more precisely the separation between the minima) and can account for a second maximum which exceeds the first.

4. Clemmow—"Edge Currents in Diffraction Theory"

Contributors: C. L. Pekeris, The Weizmann Institute. Pekeris noted the work of Schwarzschild (1912) on a treatment of diffraction by an infinite slit by iterative application of the Sommerfeld half-plane solution on the basis of multiple scattering considerations. Keller commented that Schwarzschild's work was directed toward developing the formal solution to the slit problem by the iterative procedure and establishing the convergence of that procedure, but that Schwarzschild did not attempt to evaluate the transmission coefficient nor did he consider the question of the asymptotic form of the solution for short wavelengths. Referring also to other work, Keller mentioned that of Karp who treated the problem from the point of view of scattering by the edges, his own treatment by the method of diffracted rays, the variational treatment by Levine, and the work by Fox on the acoustic pulse. Clemmow noted that he was aware of the work of Karp and of Levine which led to evaluation of the transmission cross section. He emphasized that his work was particularly concerned with the asymptotic behavior of the diffraction field, for which the only similar results he is aware of are those of Burgis discussed by Timman in another paper in this symposium. His (Clemmow's) results give the near zone region of the diffraction field in addition to the far field and the transmission coefficient; the asymptotic expression for the transmission coefficient of a slit gives values that are accurate to within one per cent for slit dimensions as small as $2\pi d/\lambda \sim 3$.

Tilston, commenting on Clemmow's formula

$$1 - \frac{1}{\sqrt{2}} e^{ika}$$

for the ratio of the electric field at the center of a circular aperture to the electric field intensity of the incident plane wave, noted that the oscillatory character of the field persists as a/λ (or ka) goes to infinity and inquired as to whether such a result is reasonable on grounds of physical considerations. Clemmow pointed out that the result is in fact borne out by the many experimental studies on diffraction by a circular aperture, and that the physical explanation is contained in the idea that the diffraction is related to sources at the edge of the aperture. The field contributed by a source on an element of arc varies

inversely as the radius of the aperture, but on the other hand, the circumference is proportional to the radius and the combination of these factors results in a resultant field of order unity.

Keller stated that the same result is obtained by the method of diffracted rays. The system of diffracted rays has a caustic which in the case of a circular aperture and normally incident plane wave reduces to the axis of the aperture. The field associated with a diffracted ray is proportional to $k^{-\frac{1}{2}}$ but the effect of a caustic for the given case is to multiply the diffracted field dependence on by $k^{\frac{1}{2}}$ resulting in an over-all contribution of order unity.

Silver added further to this discussion the remark that the result was a particular consequence of considering the incident wave to be an infinite homogeneous plane wave field. When the primary source is a point source the excitation at the edge diminishes with increasing radius, and correspondingly the oscillation of the resultant field at the center diminishes with increasing radius. In this connection he also commented on the consideration of $/\lambda \rightarrow \infty$ when the incident wave is a plane wave. His point of view is that it is meaningful to discuss only the asymptotic behavior of the field as $\lambda \rightarrow 0$, even though the scale factor of the field is a/λ, when the incident field is self a homogeneous plane wave over the infinite region; with respect to such an incident wave the limit $a \rightarrow \infty$ (with λ fixed) is not definable.

5. van der Pol—"On Discontinuous Electromagnetic Waves and the Existence of a Surface Wave"

Contributors: H. Poritsky, General Electric Company, and C. L. Pekeris. Prof. van der Pol's reference to a surface wave component in his solution evoked considerable interest. Poritsky called attention to his paper which was to be given later in the symposium in which he treats the same problem by another method. He resolves the spherical wave field into plane waves, evaluates the reflection and refraction of each plane wave, and then superposes the component plane wave fields to obtain the resultant field.

Pekeris challenged the interpretation of the term in question in van der Pol's solution as a surface wave, noting that the implied velocity of propagation of the component is higher than that of the other two terms and should correspond accordingly to a first arrival prior to the other terms at a given point of observation. There is no first arrival corresponding to that wave and, therefore, the term cannot represent a distinct surface wave component. In reply to this, Bremmer stated that Pekeris' criticism refers to the Sommerfeld surface wave but that is not what the term in the van der Pol solution represents. The van der Pol problem is that of a transmitter on the ground. The third term in the solution has the significance of a surface wave only insofar as it satisfies a two-dimensional wave equation. Any discussion of a surface wave

in the true sense must be based on a consideration of the behavior of the field as the transmitter is raised above the ground. A surface wave component should decay with increasing transmitter height. At present nothing is known about the possible behavior of van der Pol's solution for a transmitter position with respect to ground different from zero.

Keller called attention to a forthcoming paper by himself and Friedrichs in the July or August issue of the *Journal of Applied Physics*, in which they treat the problem of pulse solution for a source above the ground by the method of diffracted rays. The essence of the treatment is that there is a family of diffracted rays associated with the geometrically-refracted wave in the ground; the diffracted rays are generated at the surface. The criterion of first arrival noted by Pekeris is satisfied by the solution inasmuch as the diffracted wave cannot reach a given point on the ground until the refracted wave has reached that point.

6. Heins—"The Excitation of a Perfectly Conducting Half-Plane by a Dipole Field"

In reply to a question from Dr. Tilston, Heins amplified his presentation and pointed out that the half-plane problem is a special case of a class of problems which can be treated by the technique which he developed in this paper. The radiation from a semi-infinite circular tube can be obtained readily in a similar manner, and more generally the method permits a ready determination of solutions to the vector wave equation for a certain class of problems once the corresponding scalar type of problem has been solved. He emphasized, in commenting on a remark made by Professor Marcuvitz, that the method is exact in its technique and that the edge conditions are fully satisfied by the solution.

7. Jones—"A Critique of the Variational Method in Scattering Problems"

Contributor: R. G. Kouyoumjian, Ohio State University. Kouyoumjian commented on the parallelism between the work of Rumsey and that of Jones. He pointed out that the essence of Rumsey's method is that the stationary expression for the scattered field is a consequence of the reciprocity theorem and can be arrived at without recourse to a variational principle. Rumsey introduced the reaction principle as a basis for setting up the fundamental system of equations from which the stationary representations are obtained subsequently.

Jones noted, however, that his paper was developed from a different point of view. The basic system of equations arise without recourse to any additional physical principle. The entire development is based on a general theory of linear integral operators and follows along lines similar to those laid down by Galerkin a number of years ago in the theory of elasticity. He shows not only the equivalence between the reciprocity theorem and variational theorems but also establishes criteria for evaluating the accuracy of approximate solutions.

Dolph in commenting on Jones' results stated that they are to be expected from the general aspects of variational principles and showed how it is possible to derive the Schwinger variational principles from the usual variational principles for linear operators by requiring the variational form to be reduced to a scalar invariant representation.

8. Dolph—"The Mathematician Grapples with Linear Problems Associated with the Radiation Conditions"

Contributor: W. C. Hoffman, Rand Corporation. In response to a question from Hoffman concerning the saddle-point theory presented in the paper and its relation to saddle-point methods used in treating propagation problems, Dolph explained that the nature of the problems were not really the same. He was concerned in the present paper with the functionals for which there are various types of variational principles and the nature of the saddle-type surface which would constitute a geometrical representation of those functionals. This general problem is not directly related to the asymptotic development of integrals.

9. Keller—"Diffraction by Convex Cylinders"

Contributors: G. Toraldo de Francia, National Institute of Optics (Florence, Italy) and V. Twersky, Sylvania Electronic Defense Laboratory. It was noted by Toraldo that Keller's approach to the diffraction problem is similar in its general features to that which he had taken in some work of several years ago. In the case of diffraction of a plane wave by an aperture in a plane screen, it is found that the diffracted field given by the Kirchhoff approximation can be resolved into a system of diffracted plane waves. The nth diffracted wave obeys the relation

$$(S_n - S - V) \; x \; k = 0$$

where S_n is a unit vector defining the direction of the diffracted plane wave, S a unit vector defining the incident wave, V a vector parallel to the screen, and k a vector normal to the screen. This can be generalized to the case when either the incident wave or the screen are curved by using

$$(S_n - S - \nabla f) \, x \, N = 0$$

where N is normal to the screen and f is a suitable function defined on the screen. Diffraction can thus be treated to a first approximation by a ray method, an extension of Fermat's principle to the diffracted ray system being obtained by subjecting the stationary value of the optical path to a condition that f = constant. It then follows

that at the edge of the aperture the rays thus defined have maximum intensity when forming with the edge the same angle as the incident ray. The latter particular result was obtained by Rubinowitch in 1917.

Keller remarked that he was familiar with Rubinowitch's work and that the ideas outlined by Toraldo were contained in the general method which he (Keller) presented. The rays passing around curved bodies in his treatment are special cases of diffracted rays. It is possible to introduce a generalized Fermat's principle for diffraction, whereby a class of rays from object point to a field point is established by the seeking of a stationary path which involves an arc of the boundary of the obstacle. The total field at a point is the result of interference between

(a) direct rays from the source to the field point,
(b) reflected rays, established by the usual Fermat principle,
(c) diffracted rays, established by the conditional Fermat's principle.

The question was posed by Twersky as to how Keller's method can be applied to bodies which are not perfectly conducting and are of arbitrary shape. Keller sketched the general lines along which the analysis of such a problem would proceed, in which a system of transmitted rays is added to the field. He noted that the paper by Friedrichs and himself to which he referred in the discussion of the van der Pol paper is an illustration of a problem involving general media. Regarding the arbitrary shape question Keller remarked that it is necessary to have exact solutions for a class of problems such as the half-plane, the wedge, the circular cylinder, the sphere problems which he terms canonical problems—in order to obtain the proper distribution of energy among the systems of rays from point to point along the surface of an obstacle.

Replying to another question from Twersky as to the treatment of diffraction by a hollow cylinder having a slit in its surface, Keller stated that at the edges of the slit systems of diffracted rays are generated in accordance with the edge diffraction in the half-plane problem. Some of the rays enter into the interior where they are reflected from the walls according to ordinary laws of reflection. The multiple reflection of the rays entering into the interior develops the field inside; there is no surface wave phenomenon in the interior. The diffraction of rays at the edges of the slit together with multiple scattering of rays striking the edge from the interior make up the diffraction phenomena observed outside the cylinder.

Panel Discussion on Forward and Multiple Scattering

Chairman: J. Wiesner
Panel Members

D. Wheelon	J. Herbstreit	W. S. Ament
Twersky	J. Ortusi	H. Motz
W. Straiton	D. K. Bailey	J. B. Smyth
	N. Smith	

Wiesner: The first two questions I have are directed to Dr. Wheelon and the first one is from T. J. Carroll: "What sort of ten centimeter fields well beyond the horizon can be calculated from the Gaussian correlation functions used this morning to discuss phase variations over optical paths?"

Wheelon: I have not worked on beyond-line-of-sight problems in detail. I understand, however, that the Gaussian correlation of dielectric fluctuations does not give the right answer, either with respect to range or frequency. The exponential correlation seems to give good agreement for these same experiments. One then asks, "Why did you use the Gaussian correlation this morning?" The answer is frankly one of mathematical facility and is validated by the result that the rms phase error is substantially insensitive to the correlation model. The phase correlation between adjacent paths is sensitive to the correlation model, and provides one hope of identifying the proper correlation.

Wiesner: Maybe Carroll knows the answer. He asked the question.

Carroll: I was just afraid that the audience might draw one conclusion when you're talking about the phase variations over optical paths with the Gaussian correlation functions that you're talking about the same type of thing as is talked about in fields far beyond the horizon. There are several places where these people have tried this calculation to get the scattered fields beyond the horizon using the Gaussian correlation, which is the most logical one. And their results aren't in the right ball park, as you intimated you have heard, too.

Wheelon: Right! I do not believe, however, that one would seriously suggest evaluating scatter theory on the basis of the Gaussian correlation, since the success of scatter theory is founded on the exponential form.

Wiesner: We have a second question on the subject of this correlation function, and this one is directed to Mr. Wheelon by Dr. Norton. "Have you determined a formula for the correlation of phase received on spaced antennas for an atmospheric model in which the correlation function has a form $|e^{-|r|/l_o}$ rather than $|e^{-r^2/l_o^2}$?"

Wheelon: From simple ray theory, the phase correlation between parallel optical paths is $\rho = x K_1(x)$ where x is the ratio of the antennas spacing (d) to the turbulence scale length l_o. The exact electromagnetic method presented this morning proves very difficult when one tries to calculate with the exponential correlation. I believe that the almost complete agreement between ray theory and this new approach for the Gaussian model indicated by this morning's talk encourages one to extrapolate from ray theory to the exact answer rather freely.

Wiesner: The next question is also by Dr. Norton. "How does this compare with Herbstreit's data?"

Wheelon: I think that $\rho = x K_1(x)$ agrees quite well with the data of Herbstreit and Thompson. The phase correlation between adjacent paths for the Gaussian correlation is just $\rho = e^{-x^2}$. The large scale lengths encountered by CRPL and the 500-foot baselines; i.e., $x \simeq \frac{1}{4}$, used mean that we have not yet reached this asymptotic region of decision.

Herbstreit: We have data for separations on the order of these. As I remember it was something of that order, and in close, it seemed that the correlation function like this d/l_o factor, seemed to follow very closely an assumed curve based on an l_o of approximately 2,500 feet as I pointed out on the slide that I showed.

There is one other comment I'd like to make—that is, that the data that I showed this morning are based on a five-minute sampling period. And if other lengths of sampling periods were chosen, the form of the correlation function vs antenna spacing you would get would be different. This might be considered in terms of the, let's look at the diurnal variation only, the refractive of the phase variations. At night we'll get a very low value, in the day time we'll get a very high value. You have a very large diurnal cycle. Then you go down to a period of the order of an hour and look at the rms variations that occur inside the sampling period of an hour and you'll find that they are considerably less than would be observed over a period of a day. And that the correlation function would be less over a period of an hour than it would be over a period of a day between two adjacent paths. Then you go to a 5-minute period and again the rms value is lower and the correlation coefficient is lower, and then you can even go to smaller periods and you'll find that you still have lower rms values and lower correlation values. So actually it seems to me that there is going to be a whole spectrum of correlation functions and rms values that have to be taken into account to describe the physical processes that are involved.

Wiesner: Are there any further questions or comments from the panel or from the floor regarding this paper? If not we'll proceed on to the second. I have a question directed to Dr. Twersky from H. Meyers of the Rand Corporation. "Can you estimate the amount of signal that leaks through the grating of a parabolic reflector of a radar antenna where the grating spacing is 1/10 to 1/20 wavelength, and secondly where tolerances are about a

fiftieth wavelength; *i.e.*, how many db down is the transmitted wave from the free space wave?"

Twersky: The answer to the first part of the question is "no" and the answer to the second is "I don't know."

Wiesner: Are there any further questions or any comments regarding this paper?

Toraldo: My name is Toraldo, University of Florence. Since the question of surface waves in gratings has been raised, this reminds me of a point which I want to make—something that I had the opportunity to investigate fourteen years ago—but since I published it in an Italian journal I don't think many people have read it. I was concerned with the function, with a row of evanescent waves in all diffraction phenomena, in particular, diffraction by grating. Since evanescent waves were very well-known in total reflection, I asked myself whether there could be any connection between the two phenomena, and it appeared that there is a connection. If you consider a grating like this transparent grating and an incident wave which I will represent by ray, you have a perfect transmitted wave, and a reflected wave plus diffracted waves, and a reflected wave diffracted in all directions. Well, if you increase the angle of incidence, one wave, for example, becomes evanescent, and the other becomes evanescent at the same time; but if you have a grating on the surface separating two media of different indices, it may well be that if this is a diffracted wave, its corresponding wave is diffracted by transmission, so that when you increase the angle of incidence one wave disappears and becomes an evanescent one while the other is still a plane wave. Well, what then happens is, as soon as one wave disappears, all this energy, as far as experiment can tell, goes into the other. This is the phenomenon that I called total transmission because it is absolutely analogous to that of total refraction. I could demonstrate just by making the grating on the surface of a prism so that this wave could come out.

Twersky: I should mention that problems of this kind have been treated by Fano, who treated corrugated surface between two different media, and in fact, it was he who as far as I know first indicated specifically the role of the surface wave to account for anomalies such as those due to Wood. I would also like to say something else in connection with the question that I was asked. I stressed the literature today which dealt with the range where the wavelength is small compared to spacing. There is, of course, a much larger literature for the other range, approximations from the static case. And here the best survey for this region will be found in Marcuvitz's handbook and I don't doubt that this question that was put to me might well be resolved from the material there contained.

Wiesner: I have another written question. "How do your results on rough surfaces compare with Ament's conjectures?"

Twersky: Well, they're rather remarkably in accor with some of his conjectures, but on the whole, qui different. Perhaps I could take just a couple of minute to summarize the principal results in regard to angula and wavelength dependence of the reflection coefficien and back-scattering that I obtained. The model that treated consists of a perfectly conducting plane. To begi with, we have one arbitrary bump, arbitrary in size an dielectric properties, and we know the behavior of th bump if it were alone on the plane. And now suppose w have a plane polarized wave incident on this polarize either horizontally or vertically. And its direction is give by k_i. The specular direction with respect to the plane i specified by k_o and some arbitrary vector is given by littl vector r. These are all unit vectors. Then we extend thi to a uniformly random distribution of such bumps on plane. By uniformly random I mean that the distributio obeys ideal gas statistics. And we get the following sim ple expression for the reflection coefficient. We find tha in general, reflection coefficient can be written in terms o

the reflection impedance in the form $= \pm \left(\dfrac{1+z}{1-z} \right)^2$. Th

plus or minus here distinguishes horizontal and vertica polarization. For rough surfaces, that is the three-dimen sional problem, we find that the impedance z can be writ ten $\pi\rho/k^2$ where $k = 2\pi/\lambda$ times the scattering amplitud for one bump and is evaluated in the specular direction o reflection with respect to the plane. Here we take the do product of this with the direction of the incident e or i field for the two cases, vertical or horizontal polarization and this is divided by $\cos \alpha$ where α is the angle of in cidence. In the two-dimensional problem, that is th striated surface may be made up of arbitrary cylinders you get the same form for the reflection coefficient, which is essentially the same for the impedance except for dimensional factor, if we had a distribution of bumps, say two different kinds or some distribution in their size o dielectric properties. Then it turns out that these im pedances are all in series and we would replace z in here by, say the sum of the appropriate z's, or by an integra over this parameter whatever it is. This is the reflection coefficient, the ratio of the specular intensity to that in cident. The differential scattering cross section, that is the incoherent scattering from unit area surface, can be written in a general form.[1] The nicest thing about this general form is that you don't have to use any units. This works in all of them. Now these are the general results and the specific applications were made first to cylinders of arbitrary size and dielectric properties, and then to hemispheres. Incidentally, one gets the same form of these results not only for the electromagnetic case but also for the acoustic case. And for these arbitrary spheres or semi-cylinders, one finds that, in general, reflection co efficients near grazing incidence is of the form 1 minus

---

[1] A result from Dr. Twersky's paper was written on the board

e term of the order of beta, where beta is the horizon
gle. This result, which comes out naturally for these
bitrary size cylinders and hemispheres, is also, I believe,
general result, and I am attempting to prove this. This
ds to be the polarization. As far as the differential
attering in the back direction, we find that the horizontal
larization σ_h is of the order of β^4, and σ_v is of the order
β^2, and here we discover something rather remarkable.
ou see this result agrees with Mr. Ament's conjectures.

ment: Twersky got $1-\beta$ for his reflection coefficient.
I understand your Θ means the order of β?

wersky: I had an 0.

ment: Oh, this $0(\Theta)$. That's fine. In both polariza-
ns?

wersky: In both polarizations. But it is a different 0.
fferent consequences.

ment: That's all right. I proved that the back-scat-
er which was $p(\Theta)$, necessarily in the two-dimensional
se—went as $\Theta^2 u^2$, and therefore I don't trust your
ductions from your reflection coefficients into your
ck-scattering. Of course, you didn't do it that way.

wersky: I think you're a little confused by what I
ant by this proportionality symbol. There is no func-
nal dependence between the reflection coefficient in
e forward direction and the back scatter differential
oss section. It just turns out. It's quite clear that there
a functional dependence, however, between the reflec-
n coefficient and the total integrated cross section per
nit area of surface. That is, there is a general reflection
eorem you indicated in part. You should also have
ded the absorption per unit area of surface. And, in
ct, these results that I mentioned before that I wrote
wn. My general reflection coefficient cross section can
derived rather trivially from elementary Fresnel con-
deration and the general energy theorem.

ment: I should have said that my surface was lossless.
he surface I examined and rejected was formulated self-
onsistently and this, I believe, can be said of Twersky's
rface and Karp's surface and all the surfaces I have
ard of here, except possibly Meecham's. In that assump-
n, it is convenient—well, the surface I had was a
rallel-plate one of edges ending in the plane. I assumed
sentially that the current in each plate was identical and
epended with distance this way only according to the
ane-wave phase factor. If you then assume that the
tire scattered field is caused by currents that are pro-
uced in this manner you get bad results, and you will
nd those bad results in my paper, which is to be
ublished.

wersky: I would like to comment on this. I think
here is a fundamental fallacy in the assumption of this
ype of model for a rough surface phenomenon, in par-
cular, because it gives you completely the wrong kind
f polarization effects. I think I can show this more
mply with a picture.

Wiesner: It gives you the wrong wavelength dependence.

Twersky: I am making no great claims of the surfaces
I treat, that is, perfect planes with bumps on them are
really very reminiscent pictorially of the kind of rough
surfaces one encounters in sea-surface problems or rough
surfaces in general. However, I do believe that they show
the essential kind of incoherent scattering phenomena,
essential polarization phenomena, and also the essential
coherent phenomena. Now at this point I just want to
mention this question of polarization. We have here a
bump on a plane, an arbitrary bump, a single one with,
suppose, a plane-wave incident on it. Now for the case
of vertical polarization, we know that this problem is
equivalent to the problem of a bump of a cylinder, or
sphere, or something consisting of this plus its image,
excited by two plane waves. This is true for either polari-
zation. For the case of vertical polarization, however,
these two have the same sign. Now, as you approach
grazing incidence, the situation you come to is simply the
case of this solid object excited by a plane wave twice the
amplitude, and we know that this cannot vanish in the
forward direction. It has to remain finite because the
total scattering cross section is expressed in terms of the
forward value of the scattering amplitude. This tells us
this about vertical polarization. On the other hand for
horizontal polarization, we have the same situation essen-
tially at least in the picture, except that we have a minus
sign. And as you approach toward grazing incidence, you
end up with situation of the solid scatterer with no in-
cident wave, and the forward amplitude is as a consequence
zero. Now, the minute you take away the ground plane
here, as Ament has done, and you have a bunch of semi-
infinite plates—and I don't care what their distribution is
—you do not get this kind of an imaging effect. Now,
while it's true that in the case where you have a highly
absorbing surface you also do not get complete cancella-
tion in this limit as grazing incidence is approached, I still
believe that the perfectly conducting ground plane is really
more reminiscent of the physical problem than this. And
another thing I would like to bring out is, of course, that
you get an entirely different wavelength dependence. And
to be specific, let's consider the limiting cases—we have
hemispheres on a perfectly conducting plane. Suppose,
in one case, they are very small compared to a wavelength,
and in the other case they are very large compared to a
wavelength. Let's look at horizontal polarization first.
We see that in general the wavelength dependence of the
reflection coefficient near grazing minus unity is the same
as the back-scattering cross section near grazing and
they're both proportional to λ^{-n}. For the small spheres,
for the horizontal case we get $n = 4$, for the vertical case
we get $n = 2$. On the other hand for very large spheres,
we get $n = 2$ for horizontal, and we get zero for vertical.
For the two-dimensional problem of cylinders for horizon-
tal we get 3, for vertical we get 1, and here for large
diameters we get 2 and zero. The essential point of this
is that the horizontal polarization for either large or

small is more sensitive to wavelength—involves higher powers of wavelengths than the vertical. These results on the angular dependence that I indicated seemed to be in qualitative agreement with the summary of sea-surface data that is given in Kerr. And in a sense it's not too surprising. The two mechanisms that have been postulated for sea-surface scattering have been (1) the swells which are very large compared to wavelength, and (2) the ripples, which are very small compared to wavelength.

Wiesner: That seems to summarize this subject quite adequately. We'll turn now to the paper by Straiton and Tolbert.

Ament: I am curious to know whether you tried to explain the difference in slope of your scintillations of the vector parallel to the reflected wave vs those transverse? You may have mentioned this and I missed it.

Straiton: I said that when we first plotted those curves they had the same slope but then I realized that I hadn't taken into account the points where we got no solution and which we assumed as zero. If you throw out those points you get the same slope. If you put them in as having zero value, it distorted the slope so that I don't put any great confidence in the fact that the slopes are different or the same.

Wiesner: The next question I have is directed to Mr. Herbstreit. He is asked by T. J. Carroll to explain how he takes into account the ground reflected rays in interpreting his measured phase in the single-path experiments.

Herbstreit: The single-path experiments were conducted at 1,000 mc and both on the mountain and on the ground we used antennas which were fairly directive in the vertical plane. And the vertical angle above the horizontal from the Garden of the Gods site to the Pikes Peak site was of the order of 9° which would mean that the ground reflected signal would be of the order of 18° off of the principal axis of the antenna beam. At 18° off axis the response of the antenna is down of the order of, I think 20 or 30 db from that on the axis, which would mean that small perturbations in the phase of the ground reflected signal would not influence to any great extent the perturbations in phase on the direct path.

Wiesner: I have only one question for Dr. Ortusi and I have given it to him to read.

Ortusi: The question is the following: "Can a carcinatron quarter wave magnetron be made into a rectifier where some voltage could be used for adjusting the frequency at which rectification takes place?" I suppose it is possible, but a carcinatron is an oscillator and the noise factor will be too great. I don't know if the usual type of carcinatron can be used for a rectifier.

Wiesner: Let us turn to Dana Bailey's paper on ionospheric irregularities. I didn't receive any written questions, so will turn to the panel and see if there are any questions or comments they would like to make regarding it. Dr. Wheelon?

Wheelon: I'd like to address a question to Dr. Bailey here. Throughout the talk there was mention of going the auroral zone so as to get the effects. Maybe one of the slides ran by too fast or I missed something, but just what has the aurora got to do with this ionospheric E-layer scattering?

Bailey: Well, let me talk about that just a little bit. In the first week of observation in January, 1951, we made this very interesting observation, that during a period of moderate to intense magnetic disturbance the intensity of the scattered signal seemed to be well above what would have been expected. This was a single event, therefore not a necessarily significant one, but we thought, on the basis of the SID behavior that it might be significant, and as a result of this the research program was extended into the North. It had the dual purpose, of course, of also finding some means of communication if possible which was immune to the usual disturbances that hf experiences in the Far North or the Far South due to magnetic disturbance. Now the auroral zone, in this connection at any rate, is just merely a convenient way of referring to the circular region centered more or less on the geomagnetic pole where magnetic disturbances are most frequent and most severe. I wasn't pointing to any specific connection between this propagation and the phenomenon of aurora borealis. Though I think there are actually connections in some cases. Does that answer the question?

Wheelon: I think so. Did you notice signal enhancement when visual aurora were current?

Bailey: Now, this is a very interesting question. We made a specific search for signal enhancement during the presence of aurora in the intersecting region of our two beams. The only way we could do this, of course, was to have observers stationed at each end of the Arctic path examining periodically the direction in which the beam was located. You couldn't guarantee by merely seeing aurora in that direction that it was in the beam intersection unless the observers at both ends reported this more or less at once. Even then there was some possibility for error. Now, in this program, at least on our North-South path at Barrow, we simply didn't find any very good correlation of special events on the signal intensity recordings in the presence of aurora in the beam. We thought for awhile we might find a correlation of the Arctic sporadic E, but this, if it exists, is a poor one. There were many cases of good aurora in the beam with no unusual effects at all. Now on the other hand, this is a subject I didn't have to get into at all this morning—there are effects associated with specular reflections from the auroral shafts, as in a recent paper by Booker, Gartline, and Nichols in the *Journal of Geophysical Research.* When this goes on we get the phenomenon of very rapid fading which has been referred to in this context as sputter, which is almost without any doubt at all identical with the kinds of things Booker, Nichols, and Gartline reported from Cornell and, in fact, the same sort of thing that the

ateurs on 2 and 6 meters experience when there is aurora visible to the North. When this kind of aurora exists (it's a special kind), and also must be in a special position with respect to the path, the shafts of the aurora must be in a position so that we have an essential condition of specular reflection. But when we get this condition we then get a phenomenon of very rapid fading, a signal propagated by the aurora with very rapid fading characteristics. Now if this is stronger than the scattered signal, then this pretty well dominates what we record. Remember that we're using cw signals here, and therefore we have the summation of whatever is coming to our receivers. This phenomenon has fading rates of 200 or 300 cycles per second and in fact spreads a monochromatic spectrum line out into the spectrum. This kind of thing is very interesting and very characteristic in the auroral regions.

Wheelon: Where are the sounding devices with which you established the presence of sporadic *E* with respect to the Alaska path? Or do you actually have sounders here?

Bailey: On the Alaskan path, the best criterion for sporadic *E* on the path is the observation that the recorder simply went off scale. This has happened from time to time when the sporadic *E* intervened and gave us a signal which was very much stronger than the scattered signal. Now this could be to some extent checked by the midpoint station; there is a near midpoint station at Fairbanks, Alaska, and some further work was done on geomagnetic effects by having an ionosphere sounder at each terminal as well as an ionospheric sounder at Anchorage and one at Barrow.

Wiesner: Dr. Twersky?

Twersky: You indicated that there is some kind of controversy between two schools of thought as to the mechanism for ionospheric propagation. I wonder if you would elaborate on that a little.

Wiesner: One of the schools isn't represented here.

Bailey: Our Chairman has pointed out that one school isn't here. I'll have to do them justice; this means I'll have to bend over backwards to favor them, but never mind. On the subject of what is happening; *i.e.,* how we are getting these vhf signals, two schools of thought have developed. I think I made my point at the outset of my talk that we are still openminded on this question. As you probably realize from my hasty presentation this morning, this idea was originally sparked largely by consideration of possible reality of the tropospheric scatterer, as in Booker and Gordon. This wasn't the only thing which inspired it; however, there are independent reasons, quite independently of the possible applicability of Booker-Gordon type of thinking to the ionosphere, for supposing that the ionosphere is rough and irregularly distributed and therefore could give rise to signals that we might, for lack of a better term, call scatter. I think one has to realize in the context here that the use of the word "scatter" is used as a convenience rather than an indication of convic-

tion on the part of the people who use it that this is the correct word. It is simply a convenient word and we haven't a better one. Fading of single rays is one of the evidences we have for irregularities, re the work of Ratcliff and the group at Cavendish in the late 40's. Then there is also the ionospheric winds' experiment that leads many people to believe that there may be various kinds of turbulent motions, shears, etc., going on, particularly in the vicinity of temperature inversion in the 80-90 km region, and finally we have the much more direct evidence of what happens to long-enduring meteor trains which appear up there, which simply, if they last long enough, are observed to break up and distort quite violently. Now this is the most positive evidence of all really, because this represents motion of matter, and all these other things may be incorrectly referred to as motion of matter. Well, these are the various kinds of reasons, along with the fact that many people for a long time had observed that, with the application of higher power on oblique paths, the highest frequency from which useful signals could be gotten, from what was supposed to be sporadic *E,* often seemed to increase without any clear indication of a limit and, therefore, by going to very high powers, very directive antennas, it was possible to suppose some of the signal might be returned by something which had hitherto been called sporadic *E.* In this connection I think some of the effects which we have been getting are very closely related to what people have been calling sporadic *E.* These are the arguments in favor of the use of the word "scatter" in this connection. The other school of thought are the people, and these are represented primarily by the Stanford group in this country and the Canadian group at DRB, Forsythe, and Hogan, and McKinley at the National Research Council of Canada, who have argued that there is no need to invoke scatter at all. The entire results which we have obtained can be demonstrated, to their satisfaction at any rate, to be nothing more than the overlapping results, at a forward oblique incidence, of the scattering or reflection of signals from meteor trains, and that there are enough meteors to do this. There are many arguments that we advance against this hypothesis. They seem to continue to overlook some of these arguments, although I am well aware that they know them. The best evidence we have is the separation of the two different strata which scatter day and night. There is no optical evidence from any of the meteor work, that I am aware of, to indicate that meteors produce a significant part of their ionization much below about 90 km, with the exception, of course, of the very rare and very large ones which go all the way down to the ground. Now our daytime signals, on these southern paths at any rate, are definitely down in the vicinity of 70-80 km, and it would be very difficult by any stretch of our experimental technique or allowances for unforseen errors to get them any higher than perhaps 80 or 82. So that this poses a real problem for the people who claim that the daytime signal is due to meteors alone. You have

to be very careful not to generalize on this because there is a very sharp geomagnetic dependence which goes across the United States. When you get as far north as Boston or Ithaca things are quite different in the E layer than they are if you get down around North Carolina some-where. At night, as I think I have said, we have never denied the fact that meteors can have an important influ-ence. I just might try to use this wonderful writing machine a minute to illustrate a little more what I am talking about. Let this be a time scale here of midnight and let this be a rather unprecisely defined term which measures the amount of ionization produced in the E-layer. I won't put a letter on it or anything. Now, in general, the meteors produce. There are lots of wrinkles to this argument and I'm deliberately simplifying it quite a lot. In general, meteors are most frequent, and there-fore the largest number of meteors will go through a given area of the earth's upper atmosphere at about six o'clock in the morning and the fewest at about 1800. Now if one got slightly more esoteric in the arguments about meteors one could find that this may be a double peak at six o'clock in the morning and the minimum at 1800 may be shifted over toward nineteen or twenty. But in general, it is a very simple thing and one could represent this first term as simply a sine wave—it isn't truly a sine wave—it's more sharply peaked in the morning and really broader in the minimum. That represents the meter component coming into the atmosphere. Now the solar ultra-violet contribution to the background ionization in any region in which inhomogeneities, in this case the D or E region will always produce its maximum component about noon, which will be a function of the cosine of the zenith angle of the sun and has a function something like this: a symmetrical thing, dropping away to unmeasurable quan-tities at sunset. So that you get two terms, a sort of roughly sinusoidal component due to meteors, and one which has a simple peak at midday and fades away to unmeasurable quantities at night. In our original paper we said that the diurnal characteristics, as observed in the first few months of this experiment, could be reasonably accounted for by a combination of these two factors. In the middle of the day we felt the ultraviolet term definitely dominated, and that this was probably so well into the evening. But in the early morning hours we were at a loss to think of anything better than meteors. Since that time we have seen no reason to change this position. The height data has tended, at least, to be consistent with this view, in that in the daytime, particularly in the winter, we get heights down around 70 km. The meteor people are rather at a loss to explain the formation of any kind of scattering stratum down there resulting from meteors. On the other hand, at night, when the layer gets up to the region of 88—90 km, the meteor people are perfectly happy to accept his, and so are we, because, after all, that's what we observe. So I think this is the position, at the moment, in the con-troversy. We believe that both these factors are important,

and it's therefore of great importance to try and perfor experiments that disentangle these effects. Now the are all mid-latitude results at 50 mc that I'm talking abo now. When we go to the polar regions the picture is qui different. Up there the amplitude of that sine wave I dre awhile ago is greatly reduced, but it is still present. An the ultraviolet behavior, of course, is also very differen In the summertime the sun is up nearly 24 hours a day in fact, on some of our paths it is up for a few weeks. S that we get a great deal of ionization and direct ultraviol action there; whereas, in the winter time it is practical completely removed. Now, in spite of this, the signals the winter, with no ultraviolet at all, exhibit a diurn characteristic not unlike that which we observe in th summer except the level is much lower, and the only thin which this curve has in common with the meteoric hypo thesis is the continued presence of the minimum in th evening. But the level of this minimum is very differen Now a final word on this. In the experiments we hav performed on our paths, where we have always had highl directional antennas directed at a common volume hal way between the transmitting and receiving point, w have not been able to find, during meteor showers, an pronounced enhancement of the signal. We find a differer appearance of the signal as recorded with our time con stants, in that we find more meteors, but the average o integrated level is not significantly different durin showers. However, when you get off the great circle pat and are observing off to one side, observing the illuminate area from a point off to one side, the picture is quit different. Here, if anything, you may in some cases favo the meteors, but you certainly greatly reduce any scatte component which is fairly directionally sensitive. In thi case, if you take the integrating circuit out, you coul expect to find, I think, enhancements, significant enhance ments, during meteor showers, and this is precisely wha the Canadian group up at Ottawa has found. Just becaus we were not together on all these points, we have jus recently done a most interesting experiment. In fact th experiment is going on right now. It isn't finished, it i only started. Preliminary results, I think, can be reporte on. We've set up a second receiving station some 120-130 miles south of Sterling, Va. We chose this par-ticular location with an eye on Eshelman's theory about the probabilities of occurrence of meteor signals, in which he pointed out that we had chosen the worst possible antennas to operate our communication services with; if the signal was all meteors we had chosen to use the worst possible antennas. We knew that this wasn't so, but thought we had better do a more positive experiment to demonstrate it; and we have now performed the experiment by going down to a point which is some 7° south of Sterling, as measured at Cedar Rapids, or something like, roughly, 14° at the midpoint of the path, and erecting an absolutely identical receiving station to that at our Sterling Station, but directing the rhombic, in this case, at the midpoint area

that we are having more or less the same common volume. We now are recording, with as much care as we are able. This gets to be quite complicated. You have to compare signal generators quite carefully here, you are worrying about decibels. We now are recording continuously the signal both at Sterling and at this new place at Carrysburg, Va. with a very interesting result. Eshelman predicted that to the extent meteors were a dominant phenomenon, at least an equivalent phenomenon as a contributor to the net signal, that we ought to get more signal if we orient our two antennas, not directly toward one another, but off to the north of the geometric midpoint of the path in the northern hemisphere. This is, in effect, the exact geometry we have on this path. We've left the Cedar Rapids transmission alone, but put up a second receiving station to the south so that it is looking to the north of the center of its path. According to Eshelman we ought to get more signal when we do this. We went through his paper very carefully and put our station as nearly as we could exactly where he said it ought to be for maximum signal. The result of this experiment is very interesting. From about 11:00 at night to about 11:00 in the morning the signals down at Carrysburg are 1 or 2 db stronger than the signals at Sterling. To this extent Eshelman is confirmed. The amount is very small but they are definitely stronger. But from about 11:00 in the morning to about 10:30 or 11:00 at night the signals at Sterling are stronger. This is in the period of the day when meteors are going through their minimum. From the point of view of design of antennas for communications systems, one would design an antenna which will do the most good to you, at a time when the signals are weakest. And this is precisely what the antennas, which we have built on the existing system do. And so we feel we are vindicated and think we have also found an interesting result for Eshelman. He hasn't yet heard about this and I'm sorry he's not here. But this leaves it pretty much as we have stated all along. Meteors definitely do play a rôle in scatter propagation but they're not the whole story and at certain times they're not the dominant part of it. And with that I'll leave the controversy.

Wiesner: I think the meteors could probably be changed to cosmic dust and we would have less confusion, because it is not the large visual meteors we normally observe that are making the contribution being discussed here but rather the very small particles.

Twersky: I'm not here as an official representative of Stanford but I do have very close contact with the workers in this field. And maybe their thinking has changed somewhat since the IRE meeting, at least, I find them a lot more open-headed about this question.

Wiesner: Carroll, would you like to make a comment?

Carroll: Did he say open-headed? I just wanted to comment that perhaps Dana inadvertently gave the impression that this is only a two-way controversy, and I wanted to point out that he used the word "scatter," nearly

always, I thought, in the sense of incoherent or scatter from turbulence, whereas I believe in the ionosphere just as in the troposphere that there are adherents of coherent scatter for the phenomenon. You see, in the troposphere the controversy, which is analogous to this, is only a two-way controversy because there is no meteor effect. I should just like to point out that I drew the impression in reading over the proceedings of one of the sessions of the URSI meetings at The Hague that people like Appleton and Dellinger were attributing this phenomenon to coherent scatter from the D-layer, or sporadic D as they called it.

Wheelon: I was just going to make a wild conjecture that has occurred to me for some time. That is, perhaps this turbulence, as we know it, has to have its origin somewhere and, of course, something has to be feeding energy into the top of this decaying bubble structure and then eventually be tapped off with friction at the bottom; and perhaps, these meteors, or this meteoric dust, do in fact, as they whiz through the E-layer, provide this mechanism that initiates the turbulence, and what you are seeing at the low points is simply the decaying of the turbulence that is initiated during the meteoric period when it is highest. I guess this occurs in the night time; during the daytime it is cooling off and gradually boiling down and then comes back when the meteors return in the evening. But this is only a conjecture. It would provide a nice bridge between the two things if it were true.

Wiesner: Isn't there a wind shear layer at about this height, Dana?

Bailey: There is a temperature inversion on a supposed shear layer at about 85-90 km somewhere in there. I should point out, I think, that in defense of the meteors—I'm now arguing on their side for a moment just to show you that I'm trying to be open-minded here—you have only to introduce one additional argument here to make the meteor argument stand on a much firmer footing, and that is that somehow you have to introduce an argument which says that the electron production efficiency of a meteor is somehow related to the position of the sun in the sky, or the rate which the auroral or Arctic ionosphere is being bombarded by corpuscles. So this is the source of most of the disturbance up there. Now, if you can somehow produce an argument which says that a meteor passing through a sunlit bit of the ionosphere produces lots more electrons per unit path length than one passing through the night or one passing through a region of the Arctic ionosphere bombarded by particles, then, I think, you put the meteor people on much stronger ground. Then they can open up their arms and take on board our daytime results that our signals are much stronger in the daytime at a time when the meteoric occurrence rate doesn't fit. Now this has been suggested by ourselves in some of our reports, and I notice recently it has been suggested by Forsythe in a recent paper of his. This is an argument which it would be very interesting to know more about. We just don't have an answer on this at the moment. We

do have one very slight observational point which has not yet been confirmed and may be wrong. But when Boles was up in College, Alaska, last winter, using a radar, he was quite convinced, when they were examining the sky for meteors through a region of active aurora, that he observed many more meteors than he did at any other time. If it could be confirmed, of course, this would be again rather a direct affirmation of this point for the Arctic. It is a point which will have to be worked on some more. We just don't know the answer.

Wiesner: It's interesting to conjecture that the meteor ionization should somehow be a function of ultraviolet illumination but no one has been able to postulate a sensible mechanism to account for this to the best of my knowledge. In this connection there is one interesting additional fact, namely, that even the very late nightime-early morning signal which we are willing to attribute to cosmic dust shows a solar cycle dependence that we don't understand and which is very intriguing. This is completely not understandable from a meteor ionization point of view, unless the amount of cosmic dust is somehow correlated with the solar cycle as well.

Kearey from NEL: Isn't it possible to conjecture that during the time that the ultraviolet light is striking the ionosphere an appreciable number of molecules may be in metastable states. If this is the case the energy needed to ionize the molecule would be much less than when the sunlight is not on the ionosphere. If this were true then the meteors of smaller energy might possibly produce the ionization during the time that the atoms were in metastable states and would otherwise not.

Bailey: This is the sort of argument which I had in mind. In fact I've heard that argument made. There is another argument in connection with meteors that can be made too. We might as well have it out now since we are on the subject. That was one, to the best of my knowledge, proposed by Feinstein. He proposed that meteors being consumed by friction in going through the atmosphere leave a lot of metallic atoms behind them in the form of virtually atomic or molecular particles which therefore settle very slowly through the atmosphere. These in many cases have very low ionization potentials, and thus for periods henceforth, he predicted—and this is the thing we have never been able to confirm, partly because we have never looked for it carefully—if this were true, that after a large meteor shower such as the perseids, the gemineids, and quadrateids (which are the three biggest showers we observe by radio methods each year) for a period, depending on the rate of settling, of several days to possibly a week after the shower, our signals during the daytime, at any rate, ought to be stronger than normal due to the fact that these metallic particles are ionized so easily, and thereby increase the number of free electrons in the medium. Now this is a thing we have never looked for but here is another way that meteors can be introduced into the picure if one wishes to try to think of ways to bring them in.

Wiesner: Why don't we turn now to Dr. Motz's paper. I have several questions directed to him. First: "Could you tell us something about experimental work on the Cerenkov radiation?"

Motz: Well, as far as Cerenkov radiation is concerned there has been one paper by Danos, Cashwin, and others who had an electron beam going through a little 1-cm cavity and out of this cavity it came along. It grazed dielectric, and in this kind of experiment it is very essential that the electron beam grazes the dielectric very closely. In this experiment, powers, of the order of 10 to the minus 7 watts, were obtained at about 1-cm, I believe. That is, so far as I know, the only experiment on Cerenkov radiation, which has been performed.

Wiesner: The second question which was asked by several people is: "What is the connection of your work with that of Purcell and Abele?"

Motz: Well, there are two experiments performed on light generation. One by us. We took the assembly of magnepoles, just as I sketched in the lecture, and the magnepoles were spaced 40 millimeters from each other. There were about twenty of them and the electron beam was passed through this structure. Well, it carries out a kind of sinusoidal orbit. I say for short, in this case, the beam undulates, and we used a very high energy beam which was available at Stanford—that was a 200-million volt electron beam—and we produced visible light. In this case that is undulator radiation. Here Cerenkov radiation could not be emitted. On the other hand, Purcell and Smith, after this, did an interesting experiment where they took a grating with a grating constant of a micron or so and passed a 300 kv electron beam along the surface of this grating. They obtained a light output which, if you look at their formula for the frequency which should be received, is exactly the same as the undulator formula. However, the light emitted in this experiment is not created by the undulator effect; it is Cerenkov light. One can see when considering the electron grazing this grating that the undulating movement of the electron beam is very small compared with the amplitude of the surface charges, and can conclude that the radiation here is due to surface charges induced in the conducting grating. Therefore one has, in effect, the Cerenkov light. The formula is very similar, because in this case, by a simple Huygen's construction is does not matter whether it is actually the electron beam or the surface charges which do the radiating, and therefore the geometry of the ray optics is exactly the same, and one finds the same result as in the undulator case.

Wiesner: Are there any questions that the panel would like to direct to Prof. Motz? Are there any comments you would like to make? Are there further questions?—This gentleman back there?

Elliott: My name is Bob Elliott from Hughes. Dr. Motz, would you please comment on what factors affect the coherence of undulator radiation?

otz: Well, the coherence factor is very important. herence is a subject which has many aspects and I n't quite know what type of coherence your question ers to. One can speak about coherence in the sense t one means by it that radiation emitted by the undula- would show interference phenomena with radiation itted by another undulator or by another source which roduces radiation of a similar property. The simplest ng to say is that two undulator radiations would show erference phenomena. That is one aspect of coherence, d, of course, in that sense undulator radiation is com- ely coherent. That is the importance of the light periments, because as far as I know, this is the first periment. I mean the experiments by ourselves and rcell are the only examples of coherent light yet gen- ated. Now, there is another aspect of coherence, and t concerns the intensity in the case of millimeter ve production. The high intensities, which we have eived in our own millimeter wave experiments, were ly possible through the bunching action of the elec- ns. That's to say the electrons act coherently as they liate. Now, to understand this, one could say that e power radiated is due to the self-feed acting on the ctron. Now the self-feed will be very nearly the me in the vicinity of the electron. And so, if there is t an electron near, or if there is an electron nearby, it uld also act on the other electron, but the other electron s a self-feed, so both the self-feed and the mutual-feed ll act, and hence the power radiated will be proportional the square of the number of particles provided that the rticles are near enough, and what is near to be under- od in the following way: there will be of course, a dif- action effect; there will be consolidation of intensity if e nearby particles stretch over a full wavelength. If, wever, the bunch is small compared to a wavelength, en complete coherence will be obtained and the bunch ll act like one giant electron. Factors of a million in ensity can be obtained that way.

lliott: If you said what powers you've made by this ethod, I didn't hear it. Could you tell us what power? ou said that you had high intensity signals.

otz: Well, I wasn't asked about undulator radiation. was only asked about Cerenkov radiation so I didn't ake statements about undulator powers yet. It is hard make a statement because—

iesner: You made one. You said you had high in- nsities.

otz: It is hard to make a precise statement, but we d an energy of about a watt in the band from 1mm down perphaps a tenth of a millimeter.

iesner: This is peak—in a pulse?

otz: In a pulse. Yes.

iesner: Are there further comments or questions? ow we'll separate the men from the boys. Anyone who nts to lose his clearance will discuss the next paper. e have two questions directed to Dr. Severin. The first them: "Please indicate the reflection resulting from

each type of oblique incidence?" Shall I repeat the ques- tion for you? Did you understand it?

Severin: Oh yes. With this reflection coefficient of a dipole absorber as a function of angle of incidence—and this absorber is matched for vertical incidence—and you that means the ratio of the reflected amplitude to the incidence amplitude of nearly 0.3. That means the power ratio is nearly 10 per cent. In the next slide I could give you some idea about the possibility to match such an absorber for an angle which you wish. I mean, if you wish to have an absorber which is effective for a fixed angle of 23° in the second case here, then you have to use an absorber which is not matched for vertical incidence. I mean the dashed curve on the right, and you can achieve an absorber which is perfectly matched for the angle which you wish. Thank you.

Wiesner: I have another question: "Given an absorber coated disc, how will its diffraction pattern differ from that of a reflecting disc?"

Severin: I'm not quite sure for very high-absorbing materials, but we have made some measurements on cir- cular discs of small loss tangent and enter in diffraction; phenomena are quite nearly the same as for a completely reflecting disc. You can get each from the other by multi- plying it with the reflection factor. I think it is in a paper in 1951 in the *Zeitschrift für Angewandte Physik*.

Wiesner: I don't quite know what the author had in mind there, but possibly he'd like to amplify his quesion?

Zucker from AFCRC: I just wanted to know what would happen if the loss tangent is very large? Can you make it disappear?

Severin: We have not made investigations on diffrac- tion with very good absorbing discs, but I think we are in the beginning status with such investigations and we in- tend to measure the diffraction on a plane, on a half-plane which is perfectly absorbing. I mean—it's your question.

Zucker: Yes, the specular reflection will disappear, but what happens to the side lobe? That essentially is the question.

Wiesner: Are there further questions, either from the panel or from the floor? I thank you, Dr. Motz. There is a question here directed to Dr. deBettencourt and the panel. I suppose some of us are in the position to answer this question. "Much experimental evidence and empirical formulas are available on ground-received signal strengths in the scatter region. Is similar data available for airborne- received signal strength in the scatter region?" Now, I can start the ball rolling on this by saying that certainly there isn't as much, but there is some. There is the Air Force Cambridge Research data which was done by Mr. Rogers, and, I believe, Newman, there, in which they flew an airplane out over water distances of a couple hun- dred miles and observed the scatter signal, and when they flew close to the ground they found signals which were comparable with those that were found over land. I believe what little height exploration they did gave the kind of height dependence one would expect from a scatter

signal, but I'm not completely clear on that. Possibly Dr. Carroll can amplify this statement, with some additional data reported. Do you have any, Mr. Herbstreit?

Herbstreit: The original reason that we established our Cheyenne Mountain field station in Colorado Springs was to essentially assimulate or to find a site that looked like an aircraft site. The Cheyenne Mountain Station was essentially a fixed aircraft, at roughly 3,000 feet above the terrain. And we have conducted measurements from those, the essentially ground level, on up to the top of the mountain. These are primarily tropospheric measurements.

Guess I should have remembered that. We have to believe in reciprocity. Since you're transmitting from the height—

Ament: There were some experiments done by NRL in 1945, or 1946, written up in 1948, as I recall since I wrote them myself, dealing with the scatter propagation or long-range propagation over water, up to 200 miles at X and S band. Your altitudes flown were roughly from 100 to 1,000 feet, and the general upshot was that there was no height gain in the distant field. Once you are thirty miles or so beyond the horizon, the general fluctuating characteristics of the signal were completely different. I repeat that there is no height gain as far as we could tell in this distance.

That's a pretty dangerous statement isn't it? Because certainly when you get within line-of-sight there would be a different signal—

You're in the distance field.

Ament: Right. And the fluctuating character was completely different. I have a stack of reports that high, an anyone can write to me, or preferably to the Technical Information Officer at NRL, and those are available. I do not remember the exact number of the report. It's a very old report.

Carroll: As I think about it, since you raised the question, I find it astonishing that there are so many experiments in airplanes. Just run over the list. Gertz's first 400 mc flight has a couple of very interesting kinks in the curve. As I recall it, the only one of his airplane flights that doesn't show the scatter region is the one at 14,000 feet. I have since asked Gertz and it appears that the airplane ran out of gasoline before it got out far enough. Certainly one would have to include the NEL measurements near San Diego in this category, some of them which began, I guess about 1943, and perhaps the earliest, I think, is to go back to the March, 1933 issue of the Proceedings of the IRE, and one finds numerous airplane flights there which seem to show the effect. The unfortunate part seems to be that frequently the curves weren't carried quite far enough to show it in all the flights but it certainly is there in a large number of them twenty years ago.

Whose work was this?

Carroll: I'm thinking of both the Trevor and Carter papers, and the England, Crawford, and Mumford paper.

Wiesner: Is there any further comment on this point? Well, I think we've done well by Dr. deBettencourt. Unless there are further questions or comments from the floor I'll adjourn the meeting.

Panel Discussion on Antenna Theory and Microwave Optics

Chairman: R. C. Spencer
AF Cambridge Res. Center
Cambridge, Mass.

Members of Panel*

J. C. Simon   J. T. Bolljahn   G. Bekefi

Y. Mushiake   J. D. Kraus   R. S. Elliott

C. T. Tai   H. L. Knudsen C. L. Dolph (for E. Hallén)

Ronold King   J. Meixner   C. Toraldo di Francia

Before the meeting Prof. Toraldo announced a *Symposium on Propagation of Radio Waves in the Ionosphere* in Venice, Italy, 18-22 August, 1955. This is organized on the basis of invited papers, but anyone interested and able to attend is welcome. Further information can be obtained by writing Prof. Dario Graffi, Istituto di Mathematica de l'Universita, Bologna, Italy.

Chairman: This panel discussion is on Antenna Theory and Microwave Optics. Whereas other sessions have had their "creeping" waves, this session starts and ends with waves on wires, and sandwiched in between are a variety of other antenna and microwave problems.

1. R. King—"Theory of the Corner Driven Square-Loop Antenna"

Bolljahn: I have two questions on this paper. Prof. King showed a slide of his square-loop antenna with transmission lines connected to each of the four corners (Fig. 2). He later replaced these transmission-line elements with circuits which are not entirely equivalent (Fig. 7). It seems that his equivalent lumped circuits cannot adequately represent the transmission lines since, with some conditions of feed for which the proper symmetry does not exist, the transmission lines will be excited in an in-phase mode so that the currents at the two ends of the adjacent elements of the loop antenna will not be continuous, whereas with King's equivalent circuit we require a continuity of current. My second question is this: In breaking down the solution into component parts, Prof. King chose four different phase sequences, one in which the excitations were in phase at the four corners, a second in which there was a 90° phase lag at successive corners, another with a phase lag of 180° and a fourth of 270°. I would like to ask Prof. King if it would not be more straightforward to break the excitation into two modes, one in which the opposite corners are driven *in* phase and one in which the opposite corners are driven *out* of phase. It should be possible, by superimposing these two modes with the corresponding pair for the other

pair of corners to synthesize any combination of voltages at all four corners.

King: Dr. Bolljahn is, of course, entirely correct in his remarks with respect to this circuit. My analysis is not so general that it includes not only radiation from the loop but also radiation from all transmission lines that are attached to it. An essential restriction is that the loop is the only radiating element. Any circuit that radiates not only from the loop but also from the connecting elements cannot be handled by the methods presented here. I am afraid I was not sufficiently explicit to point this out, and Dr. Bolljahn has made a very good point. The equivalent circuit was intended to represent only the case with transmission-line distributions on the transmission lines, in which case the corner junction or coupling effect may be represented with a lumped network. But when there are wires projecting out at the corners *with radiating currents* this is a different kind of a loop. It could be analyzed, but it would be difficult.

Prof. King then illustrated his comments with diagrams, pointing out that his analysis was restricted to the case without radiating currents, that is, co-directional currents, on the transmission line. He advocated use of the two simplest modes, the zero phase sequence with all voltages equal at the four corners, and the alternating phase sequence with succeeding voltages reversed. He pointed out that the diagonal mode advocated by Bolljahn was much more complicated owing to the fact that the side is oscillating as a whole. This concluded the discussion.

2. Meixner—"The Radiation Pattern and Induced Current in a Circular Antenna with a Circular Slit"

Chairman: Three questions have been given me by J. R. Wait, of the National Bureau of Standards, for Prof. Meixner. I shall read the first one. "Would not a much better approximation be obtained if the perturbation of the current near the edge were accounted for?"

Meixner: I did not mention this in my talk except that by cutting off the current I have violated the edge condition. This idea was invented by Bouwkamp and me almost at the same time. One could of course adjust the current near the edge so that it satisfies the edge condition, but I think that the current need be corrected only within about one wavelength of the rim. I am sure that one can get better results in this manner. We are trying to verify this point, but have not yet obtained numerical results. In the case of the oblate spheroid which I mentioned, we are just now carrying through the necessary calculations.

Wait: I wish to add that in certain two-dimensional problems on which we have worked, such as ribbons, we were able to get very close agreement, as you say, by adjusting the current at the edge, to simulate the currents at the semi-infinite knife edge.[1]

Meixner: The radiation pattern is represented very well even if one violates the edge condition, but all other things, such as the radiation impedance of the antenna, cannot be approximated by such a procedure.

Chairman: The second question from Dr. Wait—"Were the exact theoretical calculations based on an oblate spheroid of vanishing or finite minor axis?"

Meixner: The calculations were carried through for an oblate spheroid with minor axis equal to zero. We could have done it, of course, for a minor axis a little larger than zero, but I think the results would not have been very much different.

Chairman: Wait's third question is: "Is it not possible that the agreement between the exact and approximate curves is only true for axially symmetric problems?"

Meixner: Well, that is a question, of course, because we have only dealt with the axially symmetric problem, but I feel quite sure that this will be true also for other distributions of sources, but I do not see a way to prove this because, at the present time, we have no rigorous numerical results for the distribution of sources which are not axially symmetrical. But this is a point which will be taken up in our future program.

Wait: I might add that one problem which can be treated in the same manner is a set of arbitrarily oriented slots on very thin elliptic cylinders, that is, one can make rigorous calculations by means of Mathieu functions and also do it with the same approximation as you did. However, if the slot is near an edge, the agreement breaks down completely. I also would like to call your attention to the work of Johnson at the Royal Aircraft Establishment in England at Farnborough, who studied slots near the edges of wings. He was able to treat rigorously the leading edge of the wing, but could only treat the trailing edge using an approximate method similar to yours, and the agreement is only fair in this case.

Meixner: Well, this is a point that must be investigated.

3. Bekefi and Bachynski—"Aberrations in Circularly Symmetric Microwave Lenses"

Chairman: This paper particularly interested me because it is a good start on the application of classical optics to microwave optical problems.

Toraldo: I would like to ask if Dr. Bekefi has considered repeating his measurements in the case where the diaphragm does not coincide with the lens edge. The reason why I think this is a good thing to do is astigmatism. As Dr. Bekefi has mentioned, astigmatism is a very bad aberration. It is bad for many reasons. One is that you cannot correct it by means of a single lens. Whenever the diaphragm coincides with the lens edge, no matter whether you are employing spherical or aspherical surfaces, or whether you are employing a single lens or a cemented doublet or triplet, the astigmatism is always equal in order of magnitude to the focal length multiplied by the square of the field angle; but if you place the diaphragm away from the lens and leave some spherical aberration, which has very little importance provided it is limited, in that case you can correct for astigmatism.

Bekefi: I have looked into Prof. Toraldo's suggestion of minimizing astigmatism by a displacement of the diaphragm, but I believe that such a system, though widely used in optics, has no practical application at microwave frequencies where the lenses are not very many wavelengths in size. By keeping the diaphragm coincident with the plane of the lens, full use can be made of the entire lens aperture and thus achieve maximum intensity in the narrow main lobe of the image. If the diaphragm is displaced, this can no longer be done because diffraction would take place not only from the hole in the aperture stop but also from the edge of the lens, and this would generally cause a serious deterioration of the image. To minimize this double diffraction effect one would be forced, either to reduce considerably the size of the diaphragm or, keeping its dimensions fixed, make an excessively large lens; neither of the two alternatives are satisfactory.

I should now like to comment briefly on Prof. Toraldo's statement that one cannot eliminate astigmatism in a two-surface lens. I believe that a fairly thick lens can be made anastigmatic. In fact Martin[2] discusses the design of an aspheric lens that would be sensibly free from spherical aberration, ordinary coma, third order astigmatism and curvature of field, provided that certain criteria as to the thickness and refractive index are obeyed.

We are planning to investigate such systems in the very near future.

Toraldo: Yes, I was referring to lenses that are more or less "thin." In that case you cannot correct for astigmatism, but if you introduce some thickness, the diaphragm cannot coincide with the lens because the lens is not a plane. . . . Toraldo then raised a question regarding the energy density diagrams which Bekefi had shown. Bekefi stated that these equal energy density contours were calculated from the Kirchhoff diffraction theory and were not densities of ray traces. He stated that the Seidel aberration theory and ray tracing theory were used only to evaluate the magnitude of the aberrations illustrated in the slides.

Chairman: We find in antenna theory that the calculation of two-dimensional integrals is very difficult.

[1] J. R. Wait and R. E. Walpole, "Calculated radiation characteristics of slots cut in metal sheets," *Can. Jour. Tech.*, vol. 33, pp. 211-227; 1955.

[2] L. C. Martin, "Wide-aperture aplanatic single lenses," *Proc. Phys. Soc.*, vol. 56, Pt. 2, pp. 104-113; 1944.

Keller: I would like to ask Dr. Bekefi if he computed any diffraction patterns himself, or if he just used the results of Nijboer for the theoretical diffraction pattern.

Bekefi: The majority of the theoretical results that were shown on the slides were obtained by Nijboer himself. However we have now a quite extensive program of computations, some of which are performed on desk calculators and others with the aid of an analog computer built by Prof. Farnell in our laboratory. The aim of the new calculations is to elucidate several questions of interest that have received as yet little or no attention. Just to mention a few; there is the study of the phase distribution and of the energy flow in the region of the focus, the study of the effects of nonuniform lens illumination in the presence of aberrations, and an investigation of caustics.

Keller: Was it necessary for you to evaluate the integral completely over the entire aperture or did it suffice to take contributions in the neighborhood of the stationary phase points?

Bekefi: Our work to date has been concerned with the effects of small aberrations. In this case the integration was carried out over the whole aperture by expanding the integrand in a series of ascending powers of the aberration coefficient and integrating term by term. If the aberrations are small, the whole aperture tends to contribute to the integral, and there are no significant areas (close to stationary phase points) which contribute to the exclusion of other regions.

I should like to add that we intend to investigate large aberrations using the analog computer I mentioned previously, and I believe that it will then be possible to analyze contributions from the various portions of the aperture. The reason is that the computer has been set up to integrate over a spiral path about the aperture center, and in the process of integration the value of the integral is recorded.

4. Elliott—"Spherical Surface-Wave Antennas"

Chairman: This paper has produced the largest number of questions to date. I am going to read the first two questions at the same time. The first is from F. J. Zucker: "How is the spherical cap antenna constructed?" The second is from J. R. Wait: "Could the author outline how the spherical surface is terminated to prevent a reflected wave on the structure?"

Elliott: The spherical caps were made out of solid aluminum stock and machined on a lathe. They are 24 inches in diameter and, in the case of the sphere constructed with a radius of two feet, the original raw stock was 3 inches thick. The surface was terminated by merely letting it stop. W. Lucke of the Stanford Research Institute showed that in the case of a corrugated plane surface, the reflection coefficient due to the end of a surface which is abruptly terminated by a ground plane, was approximately equal to the fractional amount of trapping. The corrugated surfaces that we used had about 6 per cent

trapping. This meant that the amplitude of the reflected wave is 6 per cent of the amplitude of the wave reaching the edge, assuming that the theory of Lucke can be extended. Considering the fact that about half of the power has leaked off before the wave reaches the edge, one can assume that only 4 per cent of the amplitude of the wave excited at the mouth is reflected at the termination. This results in a reflected power level of about 25 db at the feed. Therefore no attempt was made in this case to taper the depth of the teeth at the termination, although this is frequently done.

Chairman: I have some other questions on the same paper. The first of three questions turned in by Wong of the National Research Council, Canada, is: "You mention that the antenna was broadband, having a bandwidth of approximately 12 per cent. How did you measure the impedance and what was your vswr?"

Elliott: The vswr of the corrugated spherical cap antenna was measured by cutting an axial slot in the outer conductor of the coax which drove the feed. The vswr curve never exceeded 1.2 over a range from 8,400 mcs to 10,000 mcs. The pattern had only minor variations over this frequency band.

Chairman: Wong's second question is: "Is the corrugated spherical antenna limited to E_θ polarization?"

Elliott: Yes it is. Unfortunately, I did not have time to mention the dielectric-clad sphere which does not have this limitation. If a spherical surface is machined from a good conductor and then a dielectric film is bonded to this surface, one can trap waves in a manner analogous to that of the corrugated surface. Moreover, one can trap them with either polarization. If a composite dielectric film is used, one can trap both polarizations equally well and thus obtain any arbitrary polarization. We are working on this problem now and some of the experimental results will be discussed at the National Electronics Conference in Chicago by Mr. Plummer of our laboratory who is doing this work .

Chairman: Wong's third question: "How much loss in gain occurred with the null filling you accomplished?"

Elliott: I think I mentioned that with a flat corrugated surface, in an infinite ground plane, the theoretical gain was about 8½ db and that with a flat corrugated disc and no ground plane the experimental pattern gain turned out to be about 6½ db. This was reasonable, and all of the subsequent gain figures should be compared to this 6½ db. For the spherical cap with radius 6 feet, the gain dropped to 5.9 db, a loss of 0.6 db, and one may recall from the pattern that null filling was intermediate. For the more radical curvature of a 2-foot radius sphere, the gain dropped to 5.1 db. Then, when the teeth on this 2-foot radius sphere were tapered, which gave the best pattern of all, the gain rose to 5.6 db. The price paid for null filling in this case was 0.9 db with respect to the original flat prototype.

Franz: Does Dr. Elliott have a simple intelligible explanation of how surface waves are kept on the surface by rings of corrugations? Is there an easily intelligible mechanism which prevents the waves from loss by radiation?

Elliott: I think that the explanation can be offered for a flat surface and that the physical picture is similar for the cylinder or the sphere, although the mathematical functions are much more complicated. Imagine an unperturbed ground plane of a perfect conductor with a wave passing along the surface. The wave must be vertically polarized. The way a wavelength is determined is to measure the distance between two successive positions on the surface at which the current has the same instantaneous value. If this surface is corrugated by raising teeth, and there are many of these teeth per wavelength, the electric field at the top of the teeth need not be entirely normal.

For proper depths of teeth the total electric field can actually be tilted forward. The current now has to flow from the top of one tooth down across the bottom back up, down the next tooth, etc. The longitudinal distance between successive positions where the current has a common instantaneous value can be a wavelength which is not so large as in air. Therefore, with the same frequency, the phase velocity is less and the wave has been slowed down. The limit on this is that the total path length going down one tooth, across the bottom, and up to the top must not exceed half of a wavelength, for then the wave is sped up. Thus for depths of corrugations between 0 and $\lambda/4$ one can slow the wave down. The electric field now has both a vertical component and a horizontal component and these two components are in phase quadrature with each other. The component which is still normal to the surface is in phase coincidence with the magnetic field, indicating a flow of power across the surface. The component which is tangential is in phase quadrature with the magnetic field, indicating that the power that gets sent down in the space between two teeth comes back up and then progresses down between the next two teeth. I think you can see how, by varying the depths of the teeth, one can control the amount that the wave has been slowed down. The same simple picture applies in the spherical case.

In further discussion between Franz, Elliott, and Zucker it was pointed out that if the spherical cap antenna is corrugated (with or without dielectric filling), then E_θ is a permissible component of the electric field but $E_\phi = 0$. If the sphere is dielectric-clad, without corrugations, then both E_θ and E_ϕ are permissible.

5. Simon—"Application of Periodic Functions Approximation to Antenna Pattern Synthesis and Circuit Theory"

B. A. Lengyel, Hughes Aircraft: In listening to Dr. Simon's paper I was reminded of a theorem of Fejér in which he has shown that while a Fourier series of a continuous function may diverge, the arithmetic mean of the partial sums does converge. Now we have seen that Dr. Simon replaces the Fourier series by a trigonometric series and one might naturally wonder what relation the trigonometric series employed by Dr. Simon has to the arithmetic mean of the partial sums of the Fourier series.

Simon: As a matter of fact in the approximation process of bettering a Fourier sum and not a Fourier series (a Fourier sum is limited to a certain number of n terms) I used a classical linear summation process and the result of one of these processes is the Fejér sum. If I take a summing function such as $(1 - t)$ I obtain the Fejér sum but unfortunately the order of the errors is $1/n$ at best. It is possible to get a lower error by using a generalized Fejér process with a summing function $(1 - t^p)$, and, as I explained this morning, for a special case it was best to take $p = 4$.

Tai: I have a question which applies to the papers of three speakers in this session: Dr. Simon, Dr. Toraldo and Prof. Kraus. The question is that of pattern synthesis. As you know, given the power pattern, the corresponding source function is not unique. A question which can be asked is: "What further information do we need to give a practical answer to the problem? There are two kinds of problems involved; one deals with *coherent* radiation and another with *incoherent* radiation. Now for incoherent radiation what quantity do we need to investigate?"

Simon: One of the aims of my paper was to show that once you start with n distinct sources, then you have a limitation in the successive derivatives of the power pattern. It seemed to me rather important that these limitations involve only the over-all length of the antenna vs the wavelength. A second limitation is placed on the maximum error which you make when you try to approach a given function f and this was the Zamansky theorem referred to in my paper.

J. Kotik, Technical Research Group: The existence of a bound on the derivative of the power pattern determines only the length of the antenna and the wavelength. It seems to me to be completely in contradiction to the possibility of super-gaining. I wonder if any of the panel members could clear this up.

Toraldo: At first sight one might conclude that Simon's paper is in contradiction with my paper and with the ideas of people who have worked in super-gain, but this is not the case. The reason is this—Simon normalizes his pattern to V^2. If you take V^2 to be unity, you find that the angular derivatives are limited. Now what is V^2? The V^2 is the total power. It is radiated power plus reactive power. Now, we are interested only in radiated power when we are speaking of gain and this may be only a fraction of the total V^2. That is why those limitations that Simon has found may be much less effective if you have a large amount of reactive power and a small amount of radiated power. Does this answer your question?

Simon: I don't know what is the reactive power at infinity.

Toraldo: Well, I believe that the way in which you calculate your V is this. You add the amplitude, let us say, of the individual dipoles that you have in your array. Now these dipoles are assumed not to react on each other, inasmuch as one dipole does not influence the current which is feeding the other one, but this is not true for the power radiated because they are interfering with each other, so that the presence of the other dipoles may inhibit, so to speak, one dipole and impair its radiation so that the radiated power even at infinity is smaller than it would be if you could just add the powers radiated by each separately.

Simon: I think you are right when you say the interaction between the dipoles should be taken into account. I quite agree there, but when you want to look at the radiation of an antenna, you are only interested in the real part of the radiation. You can always find a distance at which there is no reactive part, if you go far enough away.

Toraldo: That is right, but please do not misunderstand me. I am not criticizing what you are saying. I only wanted to state that your paper is not at all at variance with the results of super-gain investigations.

Simon: That is right, I think. Actually I have never seen a large super-gain antenna and I wonder if it is not a mathematical trick.

Toraldo: Perhaps we may agree on that.

Kotik: It still is not entirely clear to me whether the V^2 in Simon's paper is the radiated power which one would find if one made a pattern measurement, or something else which Toraldo has called the radiated plus the reactive power. If it is only the radiated power I would like to repeat my original question, namely, that if you have a bound on the derivative of the radiated power how can you also have a super-gain antenna in which as much of the energy as you like can be put in a fixed direction, or its neighborhood?

Elliott: I wonder if, for the given length of array, whether you can get super-gaining or not depends upon n, the number of terms you are willing to take in the fabrication of this pattern. If n turns out to be greater than the number of half wavelengths in the length of the particular array, then you can begin to super-gain the pattern.

Kraus: I would like to make a practical comment about the matter of super-gain but before doing that I wonder if one of the distinctions here is between what is sometimes called *directivity* which is based only on the radiated pattern, and the *gain* of the antenna which is often defined as the signal from the antenna with respect to a reference having the same power input. In other words this latter includes the efficiency. Thus you may by super-gaining obtain an *increase* in directivity, but may at the same time get a *decrease* in gain. From a practical point of view it

might be cheaper to build a bigger antenna to get the desired increase in gain. The situation where super-gaining might be expedient is where you have an unlimited amount of money and a fixed size of aperture.

The following remarks were submitted by Dr. Simon after the symposium for inclusion in the Proceedings:

During the panel discussion Dr. Kotik asked if the limitations on $V^2(\Theta)$ derivatives, introduced by this paper, contradicts the notion of super-gain.

The answer, as it was pointed out by Prof. Toraldo, may be found in Dr. Schelkunoff's work on arrays.

In this paper it has been shown that, with

$$x = \frac{2\pi d}{\lambda} \sin \Theta$$

the power pattern at infinity $V^2(\Theta)$ for a number n of independent sources can always be written under the form of a trigonometric sum of order n, let say $P_n(x)$. This $P_n(x)$ has a 2π period in x, and represents $V^2(\Theta)$ only inside an x interval equal to

$$\left(-\frac{2\pi d}{\lambda}, + \frac{2\pi d}{\lambda} \right).$$

If $d > \frac{\lambda}{2}$, the measure of this interval is greater than the period 2π. It is no more true if $d < \frac{\lambda}{2}$. In that case, the maximum M of formulas (14), (15), and (16), which is the maximum of $P_n(x)$ over 2π, may not be the maximum of $V^2(\Theta)$ for a 2π variation of Θ, which can be called M'. As the maximum limitation of $V^2(\Theta)$ derivatives is only imposed by the factor $M \frac{a}{\lambda}$ it is possible to think that distributions may be found such as $V^2(\Theta)$ derivatives are great compared to M'. This must not be taken, of course, as a proof of the physical realizability of super-gain, but rather as the fact that the super-gain notion is not incorrect on the mathematical point of view.

6. Mushiake—"A Theoretical Analysis of the Multi-Element End-Fire Array with Particular Reference to the Yagi-Uda Antenna"

Elliott: I would like to ask Prof. Mushiake to comment on how complicated this problem gets analytically as a function of the number of elements in the end-fire array? Does he feel that if you double the number of elements that you only double the difficulty of the analysis?

Mushiake: In theory in the case of the equi-positioned n-element antenna array the number of parameters is $(3n - 1)$. If the number increases from three to four the difficulty increases I think more than twice.

Bolljahn: I would certainly agree that the difficulty would increase at a greater rate than the number of elements up to a point. If you get enough elements so that you have essentially a traveling-wave structure then the analysis is done in a quite different manner, and you think of a single mode which actually involves all of the elements in the array.

Mushiake: Yes, I quite agree with you.

7. Kraus—"Resolution, Pattern Effects, and Other Problems of Radio Telescope Antennas"

Chairman: This very interesting paper introduced the subject of radio astronomy as a whole together with its associated antenna problems. It raises a number of new questions such as, "What is it that the radio astronomers are trying to measure?" I have one question in particular I would like to ask Dr. Kraus. "When you presented typical radio telescope patterns synthesized from the convolution of typical radio astronomy objects and the antenna pattern, why did you assume the typical objects to be uniformly illuminated rectangles? Is there a reason for this other than that it is a simple assumption to make?"

Kraus: First of all, Dr. Spencer, that is a simple assumption. It is not only the simplest and easiest to apply but there are also cases where that type of distribution may approximate actual sources.

Chairman: Such as the edge of a hydrogen cloud?

Kraus: It could be a spherical surface with uniform radiation from it, or equivalent uniform distribution as we measure it with a fringed fan beam which is wide compared to the diameter of the source.

Gabor: I want to make a few remarks with which Prof. Kraus will probably agree. One of these remarks can certainly be misunderstood and that is, that the integral equation which is the convolution integral can be solved. The fact is we just haven't got the data for this solution. What we can do is this. We know the reaction of the area to a Dirac function, or delta function. And then we can, if we like, construct a reciprocal operator to that which will transform the transformed function back to the Dirac function. But this is an hypothesis. If the source happens to a single star then we get, of course, a star. But, as Toraldo di Francia has shown, it need not be a star. Even with point sources the method does not work, because the reciprocal pattern to two point sources constructed in this way will not be two delta functions. Nevertheless what is gained by this reciprocal transformation? Prof. Kraus mentioned at the beginning, *position.* Well it is probably often believed that position and resolution are connected. They are not. One can determine the position of a star if one knows that it is a point source with any accuracy limited only by noise. That is one thing. The other is if you have previous knowledge that there are two stars, or one star of a certain diameter, then again it is wrong to believe that the resolution is a limitation. It is not. That is in a paper I presented to the Manchester Symposium in April. I showed that if one uses suitable tricks, one can take the whole photon content which comes from this star and put it either into *position* or into its *diameter.* The photons are not related to information, that is, they are not specific to the data of information. It is up to the tricks of the experimenter.

This is all I wanted to say and part of it has already been said by Toraldo.

Kraus: I might add a comment that, in determining the source distribution one definitely is making use of previous information concerning what the source might be, but the limitation is more specifically that one cannot say anything about periodicidies that have a period that is less than the Rayleigh resolution, so that there are innumerable solutions which may give the same result.

Ament, Naval Research Laboratory: I have some remarks that bear strongly on both the above comments by Dr. Gabor and Prof. Kraus, and also on the information theoretic aspects of Toraldo's paper. The general idea here is that one has some prior knowledge on the radio sources. This prior knowledge is that the intensity, which is what you are measuring, is a non-negative function of position and my remarks bear on how to use this fact in transforming your integral equation into a knowledge of this intensity distribution. The idea is this. One has, in the case of a one-dimensional problem, some non-negative function $B(x)$ which is the brightness as a function of angle across the sky, and your integral equation has the form

$$\int B(x) \ K(x,y)dx \ = \ M(y).$$

(It may not be a difference kernel) where M are the measurements as a function of angle. One difficulty, of course, is that you have some noise, of unknown quantity, added in there. Another difficulty is that the necessary information in $M(y)$ may not have been acquired over a large enough range of y so that if this were a difference kernel one could not fully take the Fourier transform. Futhermore, for many kernels one has at least a discrete set of "phony" functions which can be added to this brightness distribution. That would give you no change in $M(y)$ for the finite aperture case discussed by Toraldo. There is a continuous set of such functions. Now, the question is: What can we do? There is in information theory a general idea similar to the second law of thermodynamics which may be applied here. Entropy is defined as

$$H(B) \ = \ -\int B \log B \ dx.$$

This is a quantity which has physical meaning only when $B(x)$ is non-negative. When $B(x)$ is a probability distribution in the information theoretic sense, this measures entropy or information content. Here we have a non-negative function and one wants to maximize this information, or minimize the entropy or something of that nature with respect to the integral equation as a set of constraints. The result is, essentially,

$$B(x) \ = \ \exp\left[\sum_y \lambda \ (y)K(x,y)\right],$$

and the new unknown $\lambda(y)$ must be determined through the equation

$$M(y) = \int K(x,y) \exp\left[\sum_y \lambda\,(y)K(x,y)\right]dx\,.$$

The sum on y occurs because as Toraldo pointed out there may be only a finite number of y values, or a finite number of pieces of information in your data due to a finite aperture. Obviously $B(x)$ is non-negative if λ is positive and the kernel is positive.

I do not understand this theoretically but it looks like a good way with some meaning to yield a positive answer. Computationally it might be very difficult. It turns out that this form has features that eliminate these "phony" solutions and I think this could be very important; that is, your answer obtained in this way more or less minimizes the number of bad solutions in the resulting B. This, in a way, is a quantitative use of the second law of the thermodynamics instead of the qualitative manner in which I used it in my talk yesterday.

Chairman: The discussion has progressed from a very practical talk on radio astronomy and radio astronomy antennas to fundamental problems in information theory. This I think is a good thing.

Toraldo: Perhaps I may say a few words when you come to my paper.

Chairman: Yes. This will be possible.

Gabor: I should like to answer Dr. Ament's remarks. The point is that what he is trying to give us is the theory of field theory. It has little to do with Shannon's information theory because in Shannon's information theory the means are *coding* and *matching,* neither of which is at our disposal. Coding in astronomy is done by nature herself and matching more or less in the same way. The most we can do with field theory is to filter out the noise in an optimum way, and then this criterion which you have suggested here to reduce the uncertainty to a minimum is an extremely impractical one. The only one we can practically use is Wiener's mean-square criteria. Nobody has yet tried to use it in astronomical problems because nobody has wanted to make the statistics of the things you want to observe, and that for a very good reason. Because if astronomy is a real science, then one must keep away from information theory, or that part of the information theory which has been developed by engineers for engineers. In engineering the problem is to *transmit* something which *somebody knows,* so that the other shall know it. Astronomy, being a science, seeks to *learn something* which *nobody knows,* and therefore I do not have to go further into that, but I want to say that if you want to find any new thing, then keep away from information theory.

Ament: I have kept away from information theory because I do not know anything about it. I just picked up this from hearing one talk on the subject. The trouble with the least-squares fit is that there is no guarantee in that, nor is there in taking Fourier transforms, that you will get a non-negative answer. The problem is, how do you get a non-negative answer? This is one way.

Chairman: I think there is a real problem in optimizing the design of pieces of equipment like a radio telescope, in the Wiener sense, in the light of the spectra of the things you are going to observe. If you have not observed them yet, how do you know what to use for your optimization criteria?

Simon: One could ask himself if really the mean-square criterion would be the best criterion to approach delta Dirac functions. It seems to be that the least mean-square criterion is best to approach sinusoids. What do you think, Dr. Gabor, of that question?

Gabor: I can only say that so long as you are using analysis with pencil and paper the only criteria that will permit progress is the mean-square method. If you are using computing machines, then, of course, you can do other things; you can use the maximum absolute deviation criteria or you can use the minimum——criteria.

Simon: Yes I do agree, but unfortunately it does not work when you want to get out of a filter, a delta function. You know, mathematicians have worked a lot on the fact that the least mean-square criterion was not a good criterion. That was what I tried to point out today.

8. Knudsen—"Radiation from Ring Quasi-Arrays"

Kraus: I would like to ask a question of Dr. Knudsen. You mentioned in connection with one of these ring arrays a field that was very similar to the one in the helical beam antenna. I wonder if you would just say a little more about that?

Knudsen: Yes, I think I had better use the projector. It was a case where we had a ring array of tangential antennas. We can just as well replace them with a circular wire on which there is a constant amplitude, but a progressive phase. I suppose that reminds you of one turn of the helical antenna.

Kraus: In this case the turn spacing is zero.

Knudsen: Yes, that is right. That makes a difference.

Kraus: So that it is a progressive wave in a plane and not traveling along an axis.

Knudsen: Yes. I am considering just one turn so that the pitch is zero, and therefore, the radiation pattern becomes different. In this case we get the following radiation pattern. This is the radiation pattern in the vertical plane for the five components, and this is for the Θ component, but in the helical case you get a pattern that is not symmetrical with the center of the plane.

Kraus: Yes, this is similar, but not exactly the same thing.

Knudsen: Oh no, not at all exactly.

9. Toraldo di Francia—"Directivity, Super-Gain, and Information"

Chairman: In this connection I noticed that Dr. Toraldo used the term "space frequency" or "angular frequency." Although it is common in some fields, it may not be

familiar to some of the people in the audience and I would like to have Dr. Toraldo define it.

Toraldo: Let us consider the diffraction pattern of a single point, something like this. This is a coordinate in the image plane. We may very well speak in terms of optics because, as you know, there is a perfect equivalency in these things between radio astronomy, or radio and optics. Well, you can make a Fourier analysis of this diffraction pattern. That is, the diffraction pattern can be synthesized from some sinusoidal curves having different frequencies. These are what we call "spacial" frequencies, that is, frequencies in the direction of x. Now if you have an object whose amplitude is something like this, the net result in the image is that you have to add for each point of the object a diffraction figure like the one sketched above. Mathematically what you obtain is a convolution integral. Now, there is a well-known theory in the Fourier analysis which says that if you make a Fourier analysis of this pattern, the spectrum is the *product* of the spectrum of the simple diffraction pattern and the spectrum of the object. So if the diffraction pattern does not contain any spacial frequency beyond a certain limit, the same is true of the image, and that is why the image is completely determined by a finite number of sampling points. I also, perhaps, might answer some questions by Dr. Ament and Dr. Gabor because I refrained from doing this when they were directed toward the other papers, although they did concern my paper. I only presented the case of coherent radiation where you just add the amplitudes but there is also the case of incoherent radiation. This is the one that Dr. Ament was referring to, where you have to add to the illumination which is the square of the amplitude, and you have a function which is always positive. What happens in that case, first of all, is that the spacial frequencies that are contained in the image are higher than those that are contained in the spectrum of a coherent object. This is because the spectrum of the square is the spectrum of the convolution of the function with itself, so that the frequencies are more or less doubled. But this does not say that you have more information, and this is the point that Dr. Ament raised. It is true that you have more sampling points, but you cannot choose at these sampling points a function at will, because if you choose arbitrary values for the function you may very well come out at some places where the function is negative, and that is impossible. What happens is this; you can choose the value of the function at this point, and as soon as you do this the possible variation in the limits of the function at the other points is restricted. Thus you apply constraints at the other points so that the integral which Dr. Ament showed, that gives you the entropy at a given sampling point, is limited by two limits which become narrower and narrower. When you add all the information which is contained at each sampling point, you end up with less information than you had expected. I believe this partially answers

his question. As to the question of filters that Dr. Gabor raised, I think that in the case of astronomy it is perfectly true that the objects are mostly unknown, but I would not say so about the noise. For instance, think of atmospheric noise. I think that it is not useless to analyze the spacial frequencies which are contained in this noise so as to attenuate them by means of a proper filter. In that case you can filter out your real information from your noise information perhaps.

In reply to a question, Dr. Toraldo mentioned that some of these ideas would appear in a paper of his entitled "The Capacity of an Optical Channel in the Presence of Noise." This would be published shortly in *Optica Acta* which is an international European journal of optics.

Gabor: I have two remarks concerning Dr. Toraldo's paper. The first is that he has slightly misrepresented the sampling theorem, but this is frequently misrepresented. According to the sampling theorem in the form given for the first time by Whittaker, you know the precise value of the unknown function at the sampling points, only you do not know it between. Now where are those sampling points?

Toraldo: Well generally they are equi-spaced but there is no fixed origin for the first one.

Gabor: All right, so we do not know where to put them; otherwise it would be very nice. If something interests us then we put the sampling points there, but there is no way to put it. This is always misrepresented in information theory. Now the second, I am deeply suspicious of the super-gain filters because there is a theorem with which I suppose you are very familiar concerning the reciprocity between the radius of gyration of an antenna pattern and its Fourier transform. (Gabor illustrated the following remarks with diagrams.) If you have some sort of pattern here, and some sort of x pattern here, and you take its Fourier transform, you can do something similar like this.

Simon: Dr. Gabor, could you state the limits of the transformation?

Gabor: The limits are of course infinite.

Simon: This is very important because I do not know what is infinity, I know only the limits. In many supergain realizations the way to arrive at super-gaining is to go to infinity.

Gabor: That is exactly what I am coming to. Imagine that you cut this out of paper, after first squaring it to obtain the power transform, and then you determine the radii of gyration. Let this be the angle alpha, then there is this simple theorem:

$$\Delta\alpha \cdot \Delta X \geq \lambda/2.$$

These are what are sometimes called radii of gyration or effective radii and there is the Parseval-Schwartz uncertainty relation connecting the two. Now, if one works out, as I have done, the optimum which one can get for a certain finite extent of the antenna, then one gets the

st solution. The best solution for a cylindrical antenna a half-cosine function, *cos x*, and for a rotationally symmetric antenna $J_o(r)$. Then one gets a certainty of how much it is, but one gets sidelobes that are 22 or 23 db down in this case; so I am deeply suspicious of Prof. Toraldo's result that you can put the gain up beyond any limit and put the sidelobes down below any level. But I suspect it is really the same trick as that of Woodward, who also constructed a super-gain antenna some ten years ago in the *Journal of the Electrical Engineers,* and it turns out that what he has really shown is this: You must back up your area with what amounts to a colossal dielectric baffle. That is to say, here is your antenna and you must put behind it an optically less dense medium so that there is an evanescent wave running sideways and then, of course, you can get almost anything. Dr. Simon is quite right in saying that we need not worry about directive power but we must worry about the size of the baffle, because it costs just as much as the antenna.

Chairman: I would like to comment on certain aspects of Dr. Gabor's remarks. In 1931 I published in the *Physical Review* [3,4] a general method of correcting for the smoothing effect of an apparatus. If you have an input function $G_o(x)$, where x may be distance, angle, time, et cetera, the output function $G_1(x)$ is modified or perturbed by the transfer function $F(x)$ of the apparatus. $G_1(x)$ is the convolution of $G_o(x)$ and $F(x)$.

$$G_1(x) = \int G_o(x - \alpha) F(\alpha) \, d\alpha.$$

In the case where $F(x)$ has finite moments, $\mu_n = \int x^n F(x) \, dx$, and $G_o(x)$ is continuous with continuous derivatives, $D^n G_o(x)$, it is valid to expand $G_o(x - \alpha)$ in a Taylor's series and integrate term by term, so that

$$G_1(x) = \left[\sum \frac{(-1)^n}{n!} \mu_n D^n \right] G_o(x) = P(D) \, G_o(x).$$

$P(D)$ is a generalized admittance operator. If $F(x)$ is real and positive, its second moment is positive and $G_1(x)$ has a smoothing effect on $G_o(x)$ as it "cuts the corners."

$$G_1(x) = \left[\mu_o - \mu_1 D + \frac{1}{2} \mu_2 D^2 - \cdots \right] G_o(x).$$

Thus, if μ_1 is zero the *fractional error* is proportional to

$$\frac{1}{2} \frac{\mu_2}{\mu_o} \frac{D^2 G_o(x)}{G_o(x)},$$

where μ_2/μ_o is the square of the radius of gyration of $F(x)$, which Dr. Gabor has mentioned.

Now the power diffraction pattern of a *uniformly* illuminated rectangular aperture has *infinite* second moments and this theory does not converge. However, in 1941 Norbert Wiener showed me how, by tapering the illumi-

nation over the aperture he could minimize the radius of gyration $\sqrt{\mu_2/\mu_o}$ of a diffraction pattern. For a slit he found the amplitude $f(x)$ should be *cos x* out to the first zero, and for a circular aperture it should be $J_o(r)$ out to the first zero. These are the functions just mentioned by Dr. Gabor.

I also would like to comment on the theorem quoted several times today that the spatial frequency spectrum of the image is the product of that for the object and that for the impulse response, which, in the case of incoherent points, is the antenna power pattern. The spectrum of the antenna power diffraction pattern turns out to be simply the autocorrelation function of the antenna aperture amplitude illumination. Thus, if you had a series of equally-spaced uniformly excited dipoles, whose $f(x)$ is represented by the series 1111...., the autocorrelation $A(x)$ is represented by the square of this series, 1234....321, which approximates the isosceles triangle corresponding to uniform illumination.

You can extend the idea to the autocorrelation of rectangular arrays, as I did in my Manchester paper [5], and obtain a rectangular array of practically double the number of digits in each dimension. Each row or column is again triangular in shape, which gives an idea of the relative magnitude of the space frequencies you could pick up by means of a rectangular array such as Dr. Kraus used in his radio telescope.

Now, if we could get an idea of the possible space spectra of most interest to radio astronomers (I have been told that not all radio astronomical objects are Dirac delta functions) then we might use the information in the design of a radio telescope and make a start on an optimum design. However, as the maximum number of space frequencies in any direction is simply the number of wavelengths along the maximum antenna diameter in that direction, one can only modify to some extent the intermediate space frequencies. Although as Prof. Gabor has stated, information theory cannot add anything, it certainly should be able to tell you how far you are from an optimum design in certain situations.

10. E. Hallen—"Exact Treatment of Antenna Current Wave Reflection at the End of a Tube-Shaped Cylindrical Antenna" [6]

Tilston, Sinclair Radio Laboratories: I understood from the paper that the kernel in this integral equation is e^{ikR}/R, where the R was in the form of $|R-R_0|$. I have no quarrel with this kernel. I think it is a perfectly good kernel. However, I would like to know whether there is an impression in the paper that the other kernel, namely, that due to a dipole field on the axis, was incorrect. Was this stated?

[3] R. C. Spencer, "Additional theory of the double X-ray spectrometer," *Phys. Rev.,* vol. 38, pp. 618-629; 1931.
[4] R. C. Spencer, "Some Useful Operational and Fourier Techniques," (June, 1953), to be published in Proc. of Symposium on Microwave Optics, McGill Univ., Montreal, Can.

[5] R. C. Spencer, "Antennas for Radio Astronomy," AFCRC Report TR-55-101, presented at Symposium on Astronomical Optics, Manchester, Eng.; April 17, 1955.
[6] In the absence of Prof. Hallén, selected portions of this paper had been read by Prof. C. L. Dolph.

Dolph: I am sorry I did not catch the very end of your question.

Tilston: Other people have also used a kernel e^{ikR}/R, where R is now the distance from a point on the axis to a point on the surface, and I got the distinct impression that this was considered to lead to an incorrect equation

Dolph: There seems to be, of course, a great deal of controversy in this field as to how you linearize. Now, let's go back. Are you objecting to Hallén's nonlinear integral equation, or are you objecting to the way he linearized it? I am a little confused.

Tilston: I am objecting if there is an inference that it it incorrect to use as a kernel this function e^{ikR}/R where R is the distance from a point on the axis to a point on the cylindrical surface.

Dolph: There is very definitely the statement by Hallén that this is incorrect. Would you like to have me read it?

Tilston: Yes, if you would please.

Dolph: (Quoting from Hallén's paper), "My followers have mostly written a fictitious

$$R = \sqrt{(x - \xi)^2 + a^2}$$

the distance between a point on the surface and a point on the axis, instead of $|x - \xi|$, and even used a square-root expression in the first term of eq. [5(a)] of the paper. In doing so they have not at all increased the accuracy and, in fact, have made [5(a)] unsolvable, in a strict mathematical sense."

Hallén has one other remark, which I did not quote, which might perhaps remove some of the confusion. "When there is no potential jump, or with a passive antenna and the antenna is solid, the distance

$$R = \sqrt{(x - \xi)^2 + a^2}$$

is used because one can use the vanishing electric field along the antenna axis to get an integral equation, but then there are always end surfaces." So that it would appear that the implication is just as you have interpreted it.

Tilston: I am glad you read that statement about the potential jump, because I have used this kernel in a report. However, I did use an electric field which was not discontinuous along the surface. However, there have been other papers which I still believe are valid, among them Prof. Synge's paper. He used this kernel and I believe the equation is valid for the following reason. In Prof. Synge's paper it is merely a matter of using the Lorentz reciprocity relation, and applying this relationship to the surface of the cylindrical antenna and the sphere at infinity. Both fields in this relation satisfy the radiation condition at infinity and have no singularites in the region under consideration, this region excluding the region in the interior of the antenna. Let E, H be the field which

we wish to find, and E_1, H_1 be a suitably chosen auxiliary field. If E_1, H_1 is the field of a dipole located on the axis then E_1, H_1 has no singularities in the region under consideration. Therefore, the equation as such is correct. There is a question, however, as to whether you can use this equation to get a unique answer, but the equation as such is certainly correct. The only question, as I see it could be about the uniqueness of the solution. Actually I also would like to hear Prof. King's remarks on this point.

King: From my point of view this is mostly a mathematician's argument. I think that the singularities which are discovered in the solutions and the difficulties with the integral equation in which there is a discontinuous right side and a continuous left side are largely the result of using physically fictitious generators. If a physically realizable generator or circuit is used, these difficulties do not occur. I think that the difficulty is a consequence of locating the generator at the driving point of the antenna. If the antenna is driven, for example, from a two-wire line, and an actual physically realizable generator is coupled inductively to the line, there is no problem of any singularity. Difficulties arise in the attempt to analyze an antenna with a gap which is supposed to be driven by a rotationally symmetric electric field which is maintained who knows how, and which has a current which goes off into free space at the gap. This is a physically impossible problem and it will have mathematical difficulties.

There must always be a transmission line. It may be of any type. As a simple model, consider an antenna which is connected internally to a point generator by a biconical transmission line. If the voltage across and the current in the end of this line are calculated, an impedance is obtained that involves no singularity. Now consider the problem of an antenna of finite radius which is driven by a discontinuity in scalar potential across a fictitious belt on its surface. If it is assumed that the current is in effect concentrated along the axis, there is bound to be trouble. An interesting situation arises, however, if results are calculated ignoring the mathematical singularities, and are compared with what is measured on an antenna driven by any type of transmission line. If the measured data are plotted as a function of the spacing of the line, as this is reduced progressively, the limiting values agree with the theoretical ones. In other words, the justification for a procedure which has mathematical peculiarities which no mathematician can resolve, is simply that it is a limiting case of any transmission line as its spacing is reduced to a negligible value, a physically unrealizable zero value in the limit. It appears to me that all discussion about singularities at the driving point is simply the result of trying to analyze a physically impossible setup. When you use generators which don't exist and rotationally symmetric fields maintained by nothing

all, you are bound to come out with something which does not lead to a nice mathematical form.

Dolph: I would like to say just a few words, again quoting from Hallén's paper, in answer to Dr. Tilson's question about the uniqueness. "The eq. (4) and (6), these are respectively the linearized and the non-linearized forms which we had on the slides once before) for the outgoing current waves both easily can be solved exactly. The general solution to both includes an undetermined solution to the corresponding homogeneous equation, that is, to the equation with the right-hand side put equal to zero." Then he goes on to say: "However, if we add the conditions that the solutions to (4) and (6) should have symmetry with respect to the feeding point so that the two waves going out in both directions should be equivalent, then the solution for (4) as well as (6) is unique." I should have mentioned that I was talking about the symmetric solutions here.

S. Jones, University of Manchester: The question I want to ask is about Hallén's solution of the scattering wave by a tube for which he says he has an exact solution obtained by a new method. I'm very interested in new methods, and I wonder if Prof. Dolph could possibly tell me how it differs from, and the advantages over, the standard method that has been used for this problem by Weinstein and by Pearson.

Dolph: Dr. Jones, I'm not sure I can answer your question because I do not know the papers of Weinstein or Pearson. I, myself, have not worked in this field for some time. However, I can quote precisely what Prof. Hallén says about this method. I not that Hallén does not refer to Weinstein or Pearson. There is no reference to any work connected with the Wiener-Hoph integral equation except that of the paper of Schwinger and Levine which is on the radiation of sound from the unflanged circular pipe. "The mathematical method is in its main feature that of Hopf and Wiener, (see Titchmarsh, p. 339; 1937) and it was used first on a physical tube problem (an acoustical one) by Levine and Schwinger in 1948. Our procedure is, however, entirely different from that of the latter. One of the advantages we get is that we never need a study of the field to get boundary conditions, but that these are automatically satisfied and all constants known. In fact, we need not know the formulas for the potentials and fields for solving the problem; however, we get them, as we have seen, very easily, and they are of interest in themselves." This quotation may or may not help, but it is true that he does follow through on this basis.

Jones: Thank you very much.

Tai: My comment is rather general, not specific, but it is pertinent to the problem of a cylindrical antenna. From all the works that have been going on for many years, I think I will mention that by Carson Flammer on this model

that Prof. King has sketched here. His work has not been published yet, but I think it is of interest as an exact mathematical model. He used prolate spheroidal wave functions to discuss this problem. The gap is here considered as part of the system, and we have a perfect antenna system for which we can define the impedance without ambiguity. Of course the formulation happens to be one in parameters of conductance and susceptance instead of resistance and reactance. The results for conductance compare very closely with what Stratton and Chu did many years ago, and his reactance is better in the sense that it does not have the singularity in the Stratton-Chu model. So I say I think that for the dipole problem his work seems to close a very important argument as to what's the best mathematical model. In regard to other methods, particularly those of Prof. Hallén, Prof. King, and some others, I think they are probably comparable if we don't think too seriously about the approximations that are involved in the model.

Someone from the audience mentioned that Gabor had asked Toraldo a question which he had not as yet answered.

Toraldo: The question already has been answered seven years ago but not by myself. Woodward and Lawson answered the question. It is perfectly true that when you have an aperture you can express the amplitude that you have over it by means of a Fourier transform. Each term of the Fourier series or integral (it doesn't matter what) represents a plane wave going in a given direction. Now in order to introduce exactly the amplitude here on the aperture you must have an integral that goes from $-\infty$ to $+\infty$. The variable is the sine of the angle made by the wave with the normal to the plane. This sine cannot go to infinity with a real angle, so when the sine becomes greater than unity, the plane wave becomes an evanescent wave, that is, a wave for which the E and H tangential components are in quadrature so that there is no radiation away from the screen. Now, the relation which Dr. Gabor wrote, $\triangle x \triangle \alpha$ eqauls something of the order of λ is correct, providing that you don't understand by $\triangle x$ the half-width of the x-pattern and by $\triangle \alpha$ the half-width of the α-pattern. You have to consider the whole spectrum and the whole aperture so that provided you have a diffraction pattern that has a very small central lobe and high lateral lobes, which are in the region of evanescent waves, you can have a much smaller value of this product. This settled the question, I think, from the theoretical point of view. I mentioned it in my paper and I said that I quite agree that from the practical point of view the question is really difficult, and I strongly doubt that it is feasible to obtain much higher gain than with a conventional aperture.

Bolljahn: I wish to comment on a subject akin to that of Dr. Elliott's paper. We at SRI observed the difficulty

Dr. Elliott mentioned of high sidelobe levels with an annular corrugated surface or annular trapped wave antennas, and we also considered the possibility of reducing these by placing the antenna on a nonplanar surface. In order to examine the behavior of trapped waves on certain axially-symmetric, nonplanar surfaces, we employed the sometimes useful technique of starting with a simple field solution and finding what type of boundary can support it. Unfortunately, we haven't followed this to the point that Dr. Elliott has with his sperical surface so I cannot comment on how successful these structures might be from the standpoint of sidelobe suppression.

Consider a cylindrical TM wave having a magnetic field of the form:

$$H_\phi = e^{-jhz} \, H_o^{(2)} \, (r\sqrt{k^2 - h^2}),$$

where r, ϕ, and z are the cylindrical coordinates.
This is the outward traveling type of wave with phase variation both vertically along the z axis and radially. With a solution like this it is possible to find surfaces on which the real part of $E \times H^*$ is equal to zero. At large values of r where the asymptotic form can be used for the Hankel function, these surfaces become conical, and close to the axis they deform in various ways depending on the values of the real and imaginary parts of h/k. For complex values of h, surfaces exist across which the power flow is zero but on which both tangential E and tangential H are finite. It follows that a reactive sheet could be placed on this surface having a surface reactance every where equal to the ratio of tangential E to tangential H without disturbing the solution. Hence, starting with the known solution we can find the surface for which this wave function is a solution to the traveling-wave or trapped-wave problem. It turns out that the required surface reactance function on these type surfaces is a function of radius for relatively small values of r but that it becomes a constant for $r >> \lambda$.

Chairman: Thank you, Dr. Bolljahn. I don't think that I will ask for any more questions because it is almost 5:30. I want to thank the panel and other speakers for participating in the discussions, and now we shall adjourn.

Combined Panel Session on Propagation in Doubly-Refracting Media and Future Directions for Research in Electromagnetic Wave Theory in Modern Physics

Chairman: Dr. B. Lax
Lincoln Lab.
Mass. Inst. Tech.
Lexington, Mass.

Members of Panel:

Dr. L. R. Walker
Bell Telephone Labs.
Murray Hill, N. J.
Prof. C. L. Hogan
Applied Physics Lab.
Harvard University
Cambridge, Mass.
Dr. A. A. Van Trier
Philips Research Labs.
Eindhoven, Netherlands
Prof. C. L. Pekeris
Head, Dept. Appl. Math.
Weizmann Inst. of Science
Rehovoth, Israel
Dr. G. T. Rado
Naval Research Lab.
Washington, D. C.
Dr. D. Gabor
Reader in Electronics
Imperial College for Sci. and Tech.
London, England
K. J. Button
Lincoln Lab.
Mass. Inst. Tech.
Lexington, Mass.
Prof. L. I. Schiff
Physics Dept.
Stanford University
Stanford, Calif.
Prof. H. W. Welch, Jr.
Elec. Eng. Dept.
University of Michigan
Ann Arbor, Mich.

Discussion of Papers

1. Walker and Suhl—"Propagation in Circular Waveguides Filled with Gyromagnetic Materials"

Lax: Have you compared results of perturbation theory the exact results for filled guide?

Walker: No. We have not done that except at resonance where presumably the perturbation method is not helpful because μ and K vary drastically.

Lax: Would you like to make some comments on the problem of circular waveguide partially filled with ferrite? Perhaps you have some words of encouragement for young theoretical physicists or mathematicians who might have the courage to attempt this problem.

Walker: It is an admirable and fruitful problem for some one else to tackle. There are additional parameters introduced which will make the transcendental equation more involved.

2. Hogan—"The Low Frequency Problem in the Design of Microwave Gyrators and Associated Elements"

Lax: The expression for the figure of merit for a Faraday Rotator operating above resonance indicated that it could be improved indefinitely as the field is increased. Would you discuss this further?

Hogan: One could possibly see this intuitively. The calculation are based on the assumption that the only loss mechanism is that associated with ferromagnetic resonance absorption and that the shape of the absorption line above resonance is Lorentzian. Therefore, the loss curve goes as $1/H^2$, and the differential phase shift goes as $1/H$. The loss would decrease faster than the differential phase shift so that, theoretically, the figure of merit would increase as the magnetic field is increased in the region above resonance.

Lax: I presume you are applying this to a thin pencil of ferrite?

Hogan: No. This was not applied to a pencil but calculations were based on an infinite plane wave propagating through a slab of finite thickness and infinite transverse dimensions. This is not the geometry most widely used. You can calculate by these techniques almost any geometry you wish such as the work you and R. H. Fox did at the Lincoln Laboratory in the rectangular waveguide. In most of these cases that I have done so far, I found that the particular geometry I picked appeared to be the optimum geometry. No other geometry would be better than this; none would go to lower frequencies with better characteristics. I chose this particular geometry in order to specify numbers which I think as of now are just about the rock bottom limit.

V. Twersky (Sylvania Elec. Prods., Inc.): Did you imply that you had solved the problem of a plane wave incident on a bounded slab?

Hogan: I did not worry about reflections from the slab. Is this the problem or just a plane wave being propagated through the slab? I did not try to solve an electromagnetic

wave problem. I was trying to arrive at the low frequency limitation of these devices due to the relaxation mechanism of ferromagnetic resonance.

Twersky: Then it is actually the uniform, infinite medium that you have treated.

Hogan: Yes.

Twersky: The bounded slab is still an interesting problem.

Hogan: Yes.

Lax: We have carried out calculations at Lincoln Laboratory on the bounded slab. Interesting dimensional effects can be obtained near resonance when losses are included.

G. S. Heller (Lincoln Lab., M.I.T.): Gintzberg in Russia treated the bounded slab[1] and obtained reflection and transmission coefficients. These were recently reported. Also Balzer at Lincoln Laboratory obtained the scattering matrix for a bounded slab.[2]

Lax: This is correct. However, we have now included the dependence upon the magnetic field particularly in the resonance region where the results are very interesting.

Rado: When I read Dr. Hogan's paper in the *Journal of Applied Physics*,[3] I realized that not only the magnetization but also the anisotropy must be small in the low frequency limit. One method for achieving this was to add zinc to the ferrite, and thereby decrease the effective anisotropy and hence the natural ferromagnetic resonance would be shifted to very low frequencies and gyrators, etc., would work much better. The significance of their work is that they obtained a very low anisotropy and at the same time a nonlossy dielectric constant.

Welch: Dr. Hogan stated that if a small region in the material were magnetized parallel to the applied field one situation existed and if it were magnetized perpendicular another situation occurred. He indicated that as little as a half-degree prevented the zero frequency from being reached.

Hogan: If you try to operate any microwave device at a frequency below the natural resonance of the material in its own internal anisotropy field you are licked before you start. For the device to work well the frequency of operation should be greater than the equivalent value of the anisotropy field in megacycles. I shall argue as Smit did at the NOL conference last fall. If you neglect the demagnetizing factor, for simplicity, and apply the magnetic field along the hard direction, you can saturate the crystal along the hard direction. The magnetization vector flips over into the hard direction when the applied field is exactly equal to the effective anisotropy field. Here the effective torque acting on the dipole is zero, and there

is a zero resonance frequency if you apply an rf pulse at right angles. If the magnetic field is tilted out of the hard direction by even a half-degree, then Suhl has shown that for nickel ferrite the minimum resonant frequency goes from zero to 750 mc. A field applied, in the easy direction, gives a resonant which begins at the effective crystalline anisotropy and increases with the field. In effect, if you are operating at a frequency below 2.8 megacycles times the anisotropy field, then it is impossible to obtain resonance according to Suhl. In other words, you cannot observe ferromagnetic resonance in Ferroxdure at X band.

Welch: Does the averaging method of Rado apply to the problem just described?

Rado: Yes, I think the problems are not the same because Hogan is concerned with the effective internal field. He considered extreme cases of the directions of the magnetic field with the crystalline axes. In general, there will be intermediate cases, and the average fields which I mentioned average out all of these effects. One can use these average fields for electromagnetic problems but not for predicting the actual tensor components of the material.

Welch: The constants in Maxwell's equations using the averaged field, should be measured values rather than those predicted from the solid state.

Rado: Exactly. Only under special conditions, namely, high frequencies, i.e. K band or at millimeters, can one compute the tensor components in terms of fundamental quantities such as the saturation magnetization, frequency and so on. In general, where one cannot make these approximations, at low frequencies, then all one can do theoretically is to give the form of the relation between B and H. I tried to define the conditions for which that is a tensor.

Lax: If you measure the tensor components experimentally, including the loss and the dispersion, would you put these into Maxwell's equations?

Rado: Yes. That is the point. You measure the tensor components and use them. You have to measure them for each material, if you have a complicated domain structure, cavities, inclusions, etc.

3. Van Trier—"Some Typics in the Microwave Application of Gyrotropic Media"

Heller: Dr. Van Trier, did the theoretical curve drawn along the experimental points for the pencil of ferrite in the circular guide take loss into account?

Van Trier: I did not consider losses in the theoretical calculations, but the discrepancies occur very far from resonance also. I agree with Dr. Heller that part of the discrepancies near resonance may be due to losses, but these do not account for all the differences.

Lax: If you took losses into account the actual dispersion would be less than the theoretical. Whereas for the curves shown by Dr. Van Trier, the experimental curves are above the theoretical curves, probably due to perturbation by the relatively large ferrite diameter.

[1] M. A. Gintzberg, "Dokalady," *A.N. USSR*, vol. 95, p. 753; 1954.

[2] M. Balzer, "Transmission of Plane Waves Incident Normally on Infinite Slabs of Ferrite," Lincoln Lab., Group 37 Rep.; October 28, 1952.

[3] L. G. Van Uitert, J. P. Schafer, and C. L. Hogan, *Jour. Appl. Phys.*, vol. 25, p. 925; 1954.

Van Trier: What I meant was that the difference is due to the fact that the perturbation formula giving B_+ did not apply although the radius was small. You had to go to smaller radii to get a better fit. I agree with this completely.

Hogan: These measurements were taken on a polycrystalline ferrite, which does not have a Lorentzian line shape. This is unimportant on the tail ends of the line, but near the absorption line the deviation from Lorentzian is critical in calculating the differential phase shift. Since the theoretical curves are based on a Lorentzian line, one would not expect them to fit the experimental curves for a polycrystalline material near resonance.

Lax: Measurements made by Artman and Tannenwald at the Lincoln Laboratory on polycrystalline material using cavity techniques and very small spheres, so that no dimensional effects occurred, indicated that the experimental lines fall below those drawn for a nonlossy medium. The dispersion is definitely less for magnetic fields below resonance than those for the ideal medium.

Hogan: Yes.

Lax: Dr. Van Trier, did you compare the results of your reiteration method and those obtained from second-order perturbation theory?

Van Trier: No. But I have the feeling that they should be the same in the first order.

Lax: Dr. Berk and I have done these things in a somewhat different manner. He calls his mode expansion method in which he takes into account the higher mode. I have used the second-order perturbation theory analogous to that used in quantum mechanics. It appears that you obtain expressions in the denominator which are differences of the eigenvalues as in the perturbation theory.

Van Trier: The results I have obtained are essentially the same which can be derived by Schelkunoff's method. I understand Schelkunoff's and Berk's method have much in common. I expect that they must give the same results too.

Heller: The method of Berk and Schelkunoff are almost exactly alike. It's a matter of a different type of bookkeeping for the modes.

Van Trier: That's right.

Berk: To my knowledge a mode expansion method was first given by Schelkunoff in a rather clandestine fashion which eventually became publicized. Schelkunoff used the ordinary TE-TM modes completed with the necessary third set which would account for nonvanishing divergence of the medium. I expanded the fields not in TE-TM modes—although like TE-TM modes—which individually cannot account for propagation in an empty waveguide. The advantage of this expansion was simplicity. I want to stress that the difference was one of bookkeeping, however, a rather important one since it simplified the procedure. Dr. Van Trier used another version of the same technique. Instead of using TE-TM modes he

used the ones accounting for the transverse electric field. Therefore, although there is a similarity in the methods, each has its own sphere of application.

4. Pekeris—"The Seismic Pulse, An Example of Wave Propagation in a Doubly Refracting Medium"

Twersky: The problem of the seismic pulse was pointed out earlier in the proceedings by Prof. Pekeris, van der Pol and Poritsky. During the panel discussion Keller summarized the geometrical treatment that he and Friedrichs have done. The physics of the problem is still not clear. If you could give some indication of the mechanisms and perhaps clarify your implications of the shortcomings of the geometrical approach of Keller and Friedrichs we would be obliged.

Pekeris: The physics of the situation is very simple and phenomena that I exhibited also equally simple. We expected three waves and we found three waves in the surface pulse. In the buried pulse I just indicated one phenomenon; an interplay between certain rays that come out of geometrical optics and other diffracted energies which are obtained by diffraction only. I tried to emphasize this diffracted ray which, according to geometrical optics, should carry zero energy—yet on the record it gives quite a bang. Of course, the origin of it is quite simply the fact that we are dealing here not with plane waves to which geometrical optics would apply exactly but we are dealing with spherical waves. It is this deviation between plane and spherical waves that we have to see the origin of these diffractive phenomena. The paper of Friedrichs and Keller which you mentioned, I haven't seen,—but theirs, as I understand it, is not a geometrical treatment but a wave theory treatment and it is essentially the same nature as the sort of thing that I was discussing.

Twersky: It is geometrical in the sense that Keller has been using the term throughout the session; that is, a ray treatment not restricted to the usual class of geometrically reflected rays. There are these diffracted rays which he introduces and, in principle, one can construct an exact solution to the problem in terms of these various classes of rays.

Pekeris: I like to think of diffraction as not being a part of ray theory; that is, a subject in itself.

Twersky: Then you meant the ordinary geometrical ray theory?

Pekeris: Yes, you get a reflection coefficient for the reflected ray which is 100 per cent, in ordinary plane wave theory. Therefore there is no energy to go elsewhere, and yet, by diffraction you find some energy and that is surprising and that energy is diffracted energy.

Heller: If you assume that you had such a ray as this diffracted ray and computed the travel time for that, would that give you the right travel time?

Pekeris: Yes. In fact you use the travel time to find something about the speed of the medium. It is another phase which you pick in geometrical optics. It is very

much akin to the so-called refracted ray in which the exploration industry is based; not totally, but about 25 per cent. That means hundreds of millions of dollars are spent every year on methods which are excluded by ordinary geometrical optics.

Schiff: In regard to the geometrical optics limit, one would expect this to apply when the wavelengths were short compared with the distance from the source so that one is getting essentially plane wave propagation. Now, since Dr. Pekeris was dealing with a step pulse which has all frequencies and all wavelengths present in it, one would think offhand that the geometrical optics limit might never be applicable. On the other hand, if one were to make a Fourier analysis of his displacement, in time that is, at some distance, and then look at those components which correspond to sufficiently short wavelength, then one might see how the geometrical limit is applicable.

Twersky: The geometrical optics, as used at NYU by Keller, does not refer to the usual ray optics. It is a ray procedure designed to lead ultimately to the asymptotic solutions of the wave equation and does not have the limitations of zero wavelengths of the elementary ray optics.

Schiff: I assuming the two Lame' parameters, μ and λ are equal, is this just a matter of numerical convenience to choose a particular set of values or does it simplify the analysis significantly?

Pekeris: No, it does not simplify the analysis at all. You have to assume some value and it is customary with theoretical seismologists to assume λ equal to μ.

5. Rado—"On the Electromagnetic Characterization of Ferromagnetic Media: Permeability Tensors and Spin Wave Equations"

Twersky: Dr. Rado, please amplify on triple refraction in connection with your spin wave problem.

Rado: Actually there are at least six waves instead of three. The situation is: as you remember we had the equation

$$\frac{1}{\gamma}\frac{d\vec{M}}{dt} = \vec{M} \times \vec{H} + 2A\vec{M} \times \nabla^2 M,$$

where A is the exchange constant over the saturation magnetization squared. When you combine this with the Maxwell equation, then you can see, that the $\nabla^2\vec{M}$ term will give extra waves; two extra propagation constants. The equation for the propagation constant is of the sixth order in k so there are really six waves. But we consider only three because the ones which are reflected from the back side of the metal do not really count since the metal is assumed to be very thick compared to the skin depth. Perhaps I did not emphasize that the static magnetic field is in the z direction in the plane of the sheet and that the x-component of the microwave field impinges at right angles onto the sheet. There are other solutions but they are suppressed by the waveguide usually. Of these three

waves, one is strongly attenuated and all the effects are really from two of the three. You do not really observe the three or two waves separately, but experimentally the combined effect of all the waves appears in the surface impedance, and that is the source of the effective permeability. There is then no tensor connection between B and H in this case, fundamentally because of the exchange effect. As a result of that, you can have an interesting situation appearing in the last part of the permeability, $\mu = \mu_1 - \mu_2$. This negative μ_2 ordinarily means that you have a generator. It does not mean that in this case.

Welch: Dr. Rado, you mentioned a difference in the spin wave in a material like the ferrite and in the ferromagnetic metal. Would you explain that again?

Rado: In principle, you have the spin wave effects in all ferromagnetic media. In practice, they can be completely neglected in ferrites because the conductivity is very low. If the conductivity is low, then the $\nabla^2 M$ term is very small so that you can neglect it. In metals, it had been thought also, up to our experiments, that the effect can be neglected, but it turns out that in special cases one can observe this term. I think it can only be observed in a few substances except in single crystals, where the wave propagation effects should be interesting. You can use the effect of spin waves to measure A, which is essentially the exchange integral.

Lax: You mentioned an upper frequency limit in treating the ferrite problem in relation to the permeability tensor. You said the limit would occur somewhere in the infrared. This limit should vary from material to material and depends on the Curie temperature and the anisotropy field of the ferrite.

Rado: Oh, yes indeed. That is all true. I just tried to make it simple and say that somewhere in the infrared region there has to be a limit. It will depend on the material that you use. As far as I know it has not been found experimentally, but Kaplan and Kittel have discussed it theoretically. It would be interesting to find this experimentally.

Lax: I believe they were referring to a zero field resonance in the absence of an externally applied field. They estimated that this limit might be in the region of 100 to 200 microns, or 1 to 2-tenths of a millimeter. I am not positive about these figures.

Rado: I don't remember them either. Evidently a similar thing can be carried out for the general case, that is, where there is a magnetic field.

Lax: There is an analog between the spin waves and the plasma waves and the $\nabla^2 M$ term has its equivalent in Maxwell's equation if instead of solving for the magnetic field as we do in the magnetic problems, we solve for the electric field. If we try to set up the similar problem instead of triple refraction, we get double refraction and we get a fourth-order equation for the propagation constant. Would you care to comment on this Prof. Gabor?

Gabor: Investigations of this have never been properly carried out. Everybody has considered the one-dimensional problem only, and the general equation for the anisotropic problem has not even been set up.

6. Gabor—"Plasma Oscillations"

Lax: You mentioned something about Thomson's formulation of the plasma wave problem and that it is incorrect in two regards—they ignore some of the considerations that Bohm and Gross and Vlasov use, and evidently ignore the more recent perturbation of the distribution functions that you and your students have worked out. But nevertheless, Gross has considered this problem, and he took into account the magnetic field, and made the refinement that he and Bohm considered for the simpler problem. Yet his conclusion was that the problem is so complicated, that for semi-quantitative results Thomson's approach was quite satisfactory when one is interested in learning the gross effects of the phenomena.

Gabor: It was indeed. It was wrong by a factor of 2/3. In order to follow Thomson's approach, one must be sure of a few things beforehand. First, that you have a Maxwellian distribution. Second, that you have your conditions under which the macroscopic transport equations are valid. Incidentally, this is the assumption made in most of the work of Bailey. And only these papers of Vlasov, Bohm and Gross, and Dr. Berz follow the microscopic approach. Of these I consider only the last one as free from the error of treating the fatal singularity in a cavalier fashion.

Walker: I do not understand the alarm and despondency into which you are thrown by this singularity inasmuch as this integral exists for all complex values of the frequency and would appear to have a finite limit as the frequency becomes real. It seems to me like a perfectly handy thing to have around. Why do you think it gives rise to difficulty?

Gabor: Yes, as soon as you have collisions, or as soon as you don't, you assume that the waves are damped. Indeed, that is where that integral is finite, but there is a minimum damping. That is a surprising fact which came out from Berz's work. The damping cannot be put down to zero but only to a certain finite limit. The dispersion curve which I showed was for the value of minimum damping.

Walker: I don't understand this question about damping. In the case of a continuous velocity distribution running from minus infinity to plus infinity as the Maxwellian equation would, I believe that you very frequently do not get undamped waves. If you attempt to solve the dispersion equation exactly rather than make a certain approximation to it, as Bohm and Gross, strictly speaking there is no dispersion relation in the sense that Bohm and Gross give it, because there is no solution to the equation. This is not because the dispersion equation is wrong. The solution is not right.

Gabor: Yes, in the case of the Maxwellian distribution

there are no undamped wave solutions and there are an infinity of damped solutions. What I gave was just the lower limit. In reply to Vlasov in 1946 Landau used a distribution function which extended over the complex positive plane with no poles. I cannot see what a distribution function means if you extend it over the complex plane. It happens to be right, but I do not believe them.

Walker: If the distribution were Maxwellian, there would be very little difficulty in extending that definition to complex values and to seeing that it had no poles.

Gabor: This has no physical meaning. The Maxwell distribution would have poles and essential singularities too.

Lax: In order to get the results you mention, you cannot assume that the distribution is Maxwellian; you have to make a correction. Can you say how you calculate this?

Gabor: That is a very lengthy calculation, and a part of it must be numerically calculated. The formula is very difficult and by substituting into the Vlasov equation, you can immediately verify that the equation makes sense.

Lax: Can you say something about the physical origin of this perturbation?

Gabor: The physical origin is roughly speaking associated with the exclusion of those electrons which happen to run with the wave. If you eliminate those completely, i.e., extract the poles of the integral I showed you, then there is no difficulty. Any electrons running with the wave produce a singularity. These bring to mind the famous trapped electrons which Bohm and Gross excluded from theory. The curious phenomenon arises that a plasma wave does not run through a medium like an ordinary wave which leaves the medium behind, but the plasma wave can always trap a few electrons in the wave trough and carry them along. Bohm and Gross chose to eliminate the difficulty simply by saying "take the intensity so small that there are no trapped electrons," but the mathematical difficulty does not vanish because even then you have a few electrons in the gas which run with the wave. You certainly have it if you approach the Debye wavelength because, then, the wave velocity of the plasma waves approaches the mean velocity of the electrons. This brings me back to your first question, Mr. Chairman—Nobody has treated these things properly three-dimensionally and with finite conditions. As soon as you construct reasonable three-dimensional wave packets instead of single waves, this bogey of the trapped electrons just vanishes, because if you have a finite packet then the trapped electron can slip out sideways.

Lax: I think the problem has been encountered now, to some extent experimentally, in semiconductors at low temperatures; Kittel's group in California and our group at Lincoln Laboratory have observed some effects on the cyclotron resonance at low temperatures. According to Kittel, the possibility of having the plasma bouncing back and forth as a stiff medium, from wall to wall may occur when the electrons do not freeze out as in a material like

InSb. In connection with this, we have considered a problem which looks mathematically very similar to the one Dr. Rado treated, if you put in the plasma terms and take into account the boundary conditions. Then Kittel's relations hold so long as the dimension of the medium is either smaller than an effective skin depth or smaller than a plasma wave. Then you can treat the plasma as a stiff medium.

Gabor: How small must that be?

Lax: I have not yet calculated that, but I believe it will be much less than a millimeter for the cases we considered.

Gabor: But then nobody can get dimensions less than a plasma wave because a plasma wave certainly extends to infinity, in wavelength. They have a lower limit.

Lax: The phenomenon we have observed occurred in dimensions which are larger than the ones I mentioned. I'm just placing an approximate upper limit on this, since the actual samples that have been used are of the order of a millimeter. These effects do exist.

Gabor: I am not familiar with this work but I imagine it was inspired by the Pines and Bohm theory of collective oscillations in metals.

Lax: No, I think this is a much simpler treatment. Kittel merely assumes a depolarizing effect just as he does in the ferro-magnetic case on the surface. This tacitly assumes that the plasma moves bodily back and forth with the field. I do not think you can necessarily assume this.

Gabor: This means the wavelength is long compared with the dimension.

Lax: Yes, however this has to be justified for each particular situation.

Pekeris: It has been mentioned that the distribution of the energy of the electrons in the plasma is Maxwellian. Plasma has been defined as a medium in which there are these long range coulomb forces, so long that you can never think of one electron colliding with just its neighbor; it always collides with a cloud of them. The collision with one partner is a seldom event; the common event is to collide with a whole flock of them because you are always in the vicinity of many of them. I want to ask the panel whether the Maxwell distribution law was ever proved under such conditions. I know Maxwell and Boltzmann proved their distribution law under specific conditions where they took into account only double collisions where a triple collision was already a rare event. Very likely this law holds under plasma conditions but has it ever been proved?

Gabor: It has never been proved and nobody cares. It has been, so to say, "proved" by Gibbs' methods which are not direct. The Boltzmann method has been followed up using the collision method to a few interacting partners. But approaching thermodynamical equilibrium by Gibbs' methods starting from the second principle and then looking for the equilibrium, there is no doubt that any interaction, unless it is counteracted, will present a Maxwellian distribution.

Welch: I would like to review in short some of the work that has been done on some of the properties of the plasma; the bulk properties, similiar to the work that has been done on the ferrites. This work was not included in the program of the conference because to our knowledge not much has been done in recent years. I did this work about eight or nine years ago and it has been reported. It is interesting to note some of the similarities that exist. I hope some of the people who are mathematically inclined will solve some problems which have boundary conditions. This work started with the ionosphere, propagation properties of that as a plasma. A few years ago people became interested in the magnetron which is a cylindrical structure with a cathode in the center and swarms of electrons moving around the cathode with an angular velocity. In the anode there would be some sort of a periodic structure of just a delay line. I will call it a slow-wave structure. As the anode voltage is applied to the magnetron, this swarm of electrons expands, and when the velocity approaches the velocity of the slow-wave structure, a synchronous condition is obtained, and at a little higher voltage the system starts oscillating. In plotting the resonant wavelength against anode voltage the frequency will increase as this swarm expands. During this period it may have, effectively, a negative dielectric constant, analogous to the negative permeability in a ferrite. Then when the velocity of synchronism is approached you get a rapid increase and at some anode voltage oscillation commences and the frequency comes back down. I was particularly interested in explaining this preoscillating behavior and in making measurements. The dielectric constant of such a swarm for an infinite medium has the same characteristics as the permeability of the ferrite medium, and there has been a great deal of work done on it. If you want to use a perturbation theory, ordinarily you do not know which swarms to perturb but in this connection the theory of Dr. Gabor and the theory of Ortussi, earlier in the week, is perhaps helpful.

This general class of problems include gas discharge, the electron swarms of the magnetron, and the long beam of electrons moving in the slow wave structure as in a traveling-wave tube. The paper by Dr. Lax on Wednesday morning, dealt with the plasma characteristics in a semiconductor and should have been included in this session. Another paper by Dr. Heller on Wednesday, where he considered the perturbation of the tensor rather than the perturbation of the sealar values of the permeability, is also pertinent. I believe that more experimental and theoretical work along the lines indicated here would suggest many other related problems.

Lax: Dr. Gabor, both you and Dr. Gallet mentioned that one of the mechanisms for coupling the plasma wave into the electromagnetic wave is by the use of a magnetic field in which you are propagating transversely to the magnetic field. I know that Gross has considered this problem but he did not take into account the corrections to the Max-

wellian distribution which you believe are so essential to the problem. I gather that the computation is complicated as it is. But can you at least say something about how this would affect the results predicted by Gross?

Gabor: Presumably not very much for long waves and my guess is that you should simply follow, more or less, the gross dispersion equation. The correction is much the same as that for the Bohm and Gross dispersion relation without the magnetic field. The dispersion relation does not change very much except when cutoff comes in. It is rather important and someone should work on this.

Lax: You mentioned that the plasma wave in essence dies out with one collision. Would the magnetic field affect this result?

Gabor: Presumably not. These are collisions, incidentally, which scatter over an appreciable angle. So it is not exactly the total effective cross section. And the magnetic fields could not change the results. But a few surprises may still be in store if one looks at this carefully. Incidentally, in that coupling in the magnetic field very nicely assumed by Gross, there exist no pure transverse waves, and it is quite obvious that in the presence of the magnetic field there exist no pure longitudinal waves.

Lax: Is this the case when the propagation is longitudinal along the magnetic fields?

Gabor: Yes, that is an exception to this. Then you can have pure longitudinal waves which won't radiate.

Lax: I presume this is also true in the spin waves. When you try to propagate longitudinally along the magnetic field, you do not get any coupling to the spin waves?

Rado: I think it is true that you get only double refraction instead of triple refraction. I am not sure of the details of the problem at the moment but I think it is not of real interest, because it turns out that if you have that situation it is almost impossible to think of an experiment which will give the effect. To go into this in further detail would take too long.

7. Button and Lax—"Theory of Ferrites in Rectangular Waveguide"

E. H. Turner (Bell Telephone Labs.): Were the transverse field distributions for the cases where the ferrite was away from the wall computed using the same parameters as those mentioned earlier? In this case the modes are normal waveguide modes rather than ferrite dielectric modes, are they not?

Button: The parameters used in the calculation of the case where the slab was not against the wall were the same, namely the slab thickness of one millimeter, the saturation magnetization of 3,000 gauss, a dielectric constant of 10 and the internal dc magnetic field was 1,000 oersteds. When the slab is near the wall it is a perturbed waveguide mode, but as the slab is moved away from the wall, out into the region where the electric field intensity would be high in the empty waveguide pattern, the mode turns into a ferrite dielectric mode. By this I mean, that it turns into a mode which is guided by the ferrite.

Turner: This is not, however, one of these modes which has the very peculiar behavior of transmitting in one direction.

Button: That's right.

Turner: As you let the thickness of the material go to zero, nothing peculiar happens, does it?

Lax: Let me say that there is a transition here. It turns out that when you select the parameters as we did in the first problem, that there is a definite distinction between the zero limit of the normal mode and the zero limit of what we call the ferrite mode. The ferrite dielectric mode in this case with the slab against one wall of the waveguide was compared to the guided mode of a ferrite slab bounded only on one side by a conductor. The plot of β at small thickness for these two cases was nearly identical and β increased very rapidly with decreasing thickness for the ferrite dielectric mode. Hence, in this case at zero thickness the limiting value of β did not correspond to any value of the lowest normal mode. For other parameters quite different from those chosen here β does not have these characteristics and the separation of these two types of modes is not distinct in the limit of zero thickness. This may be the case when the ferrite is moved away from the wall of the waveguide.

Turner: How do you distinguish your ferrite dielectric modes? To me the distinguishing thing was the behavior with zero ferrite thickness; the fact that you get an infinitely slow wave in the limit.

Lax: The ferrite dielectric mode is obviously correct when it is up against the wall.

Turner: I see, then this other one should not have been called a ferrite dielectric, is that proper?

Lax: This is a question of semantics, rather than in actuality. Under the conditions which you people did your experiments there does not appear to be a clear distinction. You did state that, under certain conditions the modes were perturbations of the fundamental mode and in this case you talked about the modes very much like they had properties analogous to that of the normal guide. Perhaps both points of view are correct, depending upon the selection of the parameters.

Button: I believe the term ferrite dielectric mode is fairly descriptive of what we are talking about, and that is, namely, that the wave is guided by the ferrite material, and as long as this is so the restriction on the waveguide width is not particularly important.

Turner: There is another point, and that is, if an ordinary dielectric shrinks to zero thickness, then you have a normal waveguide situation. In the case of what you have chosen to call the ferrite dielectric modes, this is certainly not the case.

Lax: If you look at the characteristic curve of this dielectric mode analogous to that of a conducting sheet against a ferrite slab, is has the same form (except for different numbers since you are including the permea-

bility properties) as the dielectric modes discussed by Schelkunoff.[4] It has the same character.

Turner: It has an hyperbolic dependence.

Lax: No, I mean the beta as a function of thickness. This is how we identify it.

Turner: Does not the case where you claim propagation in one direction only for a two-terminal pair system violate the conservation of energy and/or the second law of thermodynamics? In other words, it presumably violates the first and second law.

Button: First, I believe there are no propagating TE modes missing from the diagrams that I have shown on the screen. I believe that all of the TE modes are there. That is, if one or more modes were missing then you would not be able to postulate a one-way transmission line or anything like it. Secondly, in applying the second law of thermodynamics one must have an appropriate model which includes this element in order to discuss the possibility of the violation. Apparently no convincing models have been proposed. I do not say that a model cannot be drawn which demonstrates the violation.

Lax: The only piece of information which is concrete is the analysis that we have. And this we have checked and rechecked, and we believe it. It seems to indicate the existence of this paradox. Now if it is possible to describe this element as a two-terminal network (which may be questionable) and one works out the scattering matrix for such a system one runs into a contradiction. So this appears, on the surface, without very deep examination, that the device may violate the first law of thermodynamics. The other argument is a phenomenological one, which may be even more questionable in the versions that I have heard, since it brings in a number of artificial filters and black bodies and it is not clear what the spectrum of the black body is. The claim is that if you place a black body on one side of the device then you are going to pour energy from low temperatures to high temperatures without doing work. We recognize the paradox and more research done on this problem probably will resolve it.

Button: These curves of β versus δ that were shown here on the screen are correct. Whether the one-way transmission line works or does not work is not a question of the validity of these curves.

Walker: I would certainly like to agree with you. You have the only real roots; I think we will all agree.

Heller: Most of the thermodynamic arguments that I have heard regarding this device assume a very ideal situation which starts out with the proposition that energy propagates in one direction and not in the other. If you are going to start out with that assumption then, in fact, you have started out by violating it. I contend that the premises involved do not apply to this problem at all. In the case where you have several modes you can not think of this as a two-pole device. You must have a pole for each mode; also if you put an arbitrary black body into a waveguide you should expand the field in terms of all the modes. This also means all the cutoff modes. One way to get around this thermodynamic argument is to say that this thing is based on loss. But then the argument is: "But you never introduce loss into the equations?" Certainly, loss can be included in the cutoff modes and that dispels that kind of thermodynamic argument.

Walker: One should not start to talk too much about the second law here because so few people really understand it well enough to be sure of what they are saying, speaking for myself, anyway. It seems to me that there is a difficulty in this problem in any case which is simply connected with the first law. As Dr. Lax said, you might assume that this is simply a two-port device. In other words if you put empty waveguide at each end which will propagate only the dominant mode, and if you went far enough away at each end then you could suppose that you had only the amplitudes of the ingoing and outgoing dominant modes. The condition that the device is lossless insures you that if you look in at one end with the far end of the waveguide terminated, you will see a certain reflection coefficient. Then if you go to the far end of the device, terminating the near end, you will see the same reflection coefficient, differing only by a phase factor. This means that the device behaves in a sort of quasi-reciprocal way. It does not show signs of transmitting power effectively in one way and not in the other. It seems to me that the real paradox is, how does this actually come about in practice, if in the section containing the ferrite there is only one genuinely propagating mode? Does the possibility exist that you could actually fit boundary conditions and thus excite various amplitudes of cutoff modes which would bring everything out in accordance with the first law? Now you might imagine for a certain fixed length that you had done this. Then, if you make the ferrite section fifteen times longer, you will either have to assume that the amplitudes of the cutoff modes that are excited in order to do the job are simply enormous, or else the amount of the propagating mode that is excited is extremely small. In any case no one has a clear picture of the disposition of the fields inside the ferrite which will satisfy the first law. That is where the real problem is apart from any questions about the second law.

Most people, when they are looking for modes of propagation, simply look for real roots, and if the root leaves for the complex plane, it is ignored from then on. I thought the possibility might exist that the two genuine waveguide modes which have now disappeared in this problem under these conditions of magnetic field had not gone very far into the complex plane; that they were not very strongly cut off. So I looked for these roots, and it turns out that this is decidedly not so. They have about

[4] G. C. Southworth, "Principles and Applications of Waveguide Transmission," D. Van Nostrand Co., Inc., New York, N. Y., p. 129; 1950

one neper per wavelength. So they are really quite strongly cut off. There is no plausibility argument of that kind.

Lax: Mr. Button has also found this complex mode, I do not know just what numbers he obtained but perhaps he might tell us.

Button: One can almost see this intuitively. As I pointed out, the TE waveguide-type mode is being forced into an effectively narrower waveguide and if this process is continuous, then the cutoff that occurs due to this phenomenon is just as strong at that in which you actually narrow the waveguide down. One knows in that case that the wave is attenuated very sharply thereafter.

Gabor: It is generally impossible to break down the second law of thermodynamics with a lossless device. In a lossless device, the second derivative with respect to time will come in. In other words, the system will be reversible. It has been shown by many speculators a long time back that it is perfectly impossible to take any machine or demon and make it break the second law.

Lax: We do not disagree. The question is: can you apply the second law of thermodynamics to this system as it actually operates under the chosen conditions in a microwave circuit?

Gabor: Certainly not.

Lax: We are saying that you can build a nonreciprocal device using a four-terminal network.

Gabor: That's right.

Lax: Perhaps the problem resides in: Is this a true two-terminal network, or is it, in essence, a four-terminal network? You have four modes in this device; we described four modes. Does each mode represent, intrinsically, a terminal?

Walker: I do not understand your last remarks at all. What do you mean by saying that you have four modes? You have an infinite number of modes.

Lax: I say four modes in the sense that the field configuration for each of them is different. You have two of what we might call ferrite dielectric modes in the two directions of propagation. And then we have two modes which are cut off which are more like the distortion of the normal waveguide modes.

Walker: Yes, but I don't see why you should deny to those modes which have been cut off since birth, their natural rights.

Lax: Oh, but we are not. We are saying that they are there. They have values of β in the complex plane.

Walker: Then I don't see why you should mention four modes or any finite number of modes particularly. And why do you raise the question of this being a two-terminal or four-terminal?

Lax: There are higher modes too, which have a complex propagation constant. However, the eigenvalues of the higher modes are of such magnitude in the complex plane that in any perturbation calculation involving interaction of these modes only the four I mentioned need be considered.

Walker: Would you say that there was any doubt, for instance, about the application of this theorem using the scattering matrix? I mean, you would agree that you could get far enough away so that there was nothing but the dominant modes in the waveguide.

Lax: Yes, there may be some doubt.

Walker: One possibility presumably might be that this device works exactly the way you say it works but that if you take a substantial length of this thing that the amplitudes of the cutoff modes which are demanded in order to satisfy the boundary conditions would just make it necessary to take into account the actual loss in the material much sooner than you would ever have thought.

Gabor: Just a little imaginary experiment: Imagine that you take your device and put on one side a body at a certain temperature T and on the other side a 100 per cent reflecting mirror. Now all the modes will get the energy kT. Now take the mirror away and put a body of the same temperature there and absolutely nothing will happen. It is already in equilibrium. You can't break down the second principle with this sort of thing.

Hogan: I think that we are all agreed that there is some kind of a paradox which none of us really understands.

Turner: The longitudinal component of the rf magnetic field does not equal to zero only because the solutions with the hyperbolic dependence in air-filled regions are chosen. The inference that I drew from this was that previous publications[5] were incorrect because they stated that you could have a longitudinal h equal to zero. In this new work you are reporting now, only one set of parameters was chosen and hence too many conclusions cannot be drawn from the results.

Lax: We made one assumption which need not necessarily be made to make the field displacement devices work. We assumed that it is desirable to choose only this dielectric mode, in other words, cut off the other modes, for optimum design of such device. Under these conditions, as you point out, hy does not go to zero. However, if you do not assume this, you have these other modes in the device. How do you suppress them for optimum operation of these new devices?

Turner: Experimentally, our results to date have not indicated that the optimum case is one where you use what you term the ferrite dielectric modes. For example, in the field displacement isolator, in my paper last Wednesday, I believe anybody who heard it would have realized that it was a normal waveguide mode that was being used. And, from the experiments which various of us have done at Bell Laboratory at Holmdel, I don't think there is any doubt that the most practical cases have been ones where we have not had very large ferrite fillings of the waveguide and in general we have not been operating in this ferrite dielectric region. That is, they are normal waveguide modes for most practical situations.

5 B. Lax, K. J. Button, and L. M. Roth, *Jour. Appl. Phys.*, vol 25, p. 1413; 1413; 1954.

Button: How can you be sure which mode you are propagating unless you have calculated all modes exactly using the actual parameters of your experiments? I have not done so because I could not tell the Fox, Miller, and Weiss paper what these were.

Turner: In specific instances we certainly have. This is precisely what I did for the paper that I gave Wednesday. We find an electric field null and the experiments certainly justify it, I think. We were using normal waveguide modes. For the normal waveguide mode, I find an electric field null for the dominant mode, a null which is in the interior of the waveguide. We have the behavior of this mode as a function of magnetic field and ferrite thickness, and so forth, and the experiments agree very nicely. So to my mind, there is no question.

Discussion of Future Directions for Research in Electromagnetic Wave Theory in Modern Physics

Lax: At this time I would like to ask Prof. Schiff to make some remarks on the last topic on our discussion panel.

Schiff: I realize that this is an extraneous topic and would only like to point out one way in which I think that some of the work that could be done in electromagnetic wave theory could be of value and interest to current research in modern physics. In talking about modern physics, I will arbitrarily restrict it to atomic or nuclear physics. This, of course, leaves out a lot of other interesting topics.

In the first place, one might ask: "Under what circumstances can classical electromagnetic theory be used in connection with atomic or nuclear physics?" In order to answer this, one has to see when the scattering of electromagnetic waves, for example, by an atom or by a nucleus, can be treated classically. It is evident that if the wavelength is sufficiently long this can be done. If the wavelength becomes comparable with the size of the system, then one has to inquire whether the energy content of a single quantum of this radiation is comparable to the energy content of the object which is being scattered. An easy way to decide this is to write down an expression for the energy content per unit volume of each particular type of matter. In the case of normal atomic matter this is of the order of magnitude of one electron volt per cubic $\mathring{A}$. This is about the amount of energy that one finds in the electronic structure of atoms or solids. In the case of nuclear matter the figure is approximately 1 Mev. in the cubic region whose dimensions are 10^{-13} cubic cm. If I call the first one the atomic energy content and the second the nuclear, the second one turns out to be about 10^{21} bigger than the first

Now, write down the ratio of the energy contained in the volumes whose dimensions are L (which is ε L^3 to the quantum energy of an electromagnetic photon or quantum which has the wavelength of order L. By wavelength I will use the symbol λ which is the ordinary wavelength divided by 2π. This is $\hbar\, c/L$ where $\hbar$ is Planck's constant divided by 2π and c is the speed of light. This ratio gives a measure of the energy content to the quantum energy. If R is large compared to one then the classical theory can be used for electromagnetic scattering even for short wavelengths because the wave will always have a small quantum energy. If, on the other hand, the ratio is less than or of order one, then quantum theory must be used except for long waves, and then one loses all the interest of the diffraction theory which is taking up so much of the attention of electromagnetic scattering problems at the present time.

Well if one puts in some numbers, one finds that setting R equal to one, which is kind of a division point, the length which comes out in the atomic case is approximately 10 $\mathring{A}$ units and the length that comes out in the nuclear case is approximately 10^{-12}cm. In both cases these are larger than the objects that we are interested in. The length in the atomic case is around 1 or 2 $\mathring{A}$ and in the nuclear case is around a few times 10^{-13}cm. So one is forced to conclude that the classical theory is not directly useful in connection with electromagnetic scattering problems as applied to atoms and nuclei.

There might, however, be exception in the case of scattering of electromagnetic waves from large molecules, which could have dimensions somewhat larger than 10 $\mathring{A}$ units. And this is one of the items that I would like to set aside as of possible future interest to electromagnetic scattering theory. Micro-crystals are another possibility. The wavelengths would then be in the soft X ray region of the order of 100 $\mathring{A}$ or so and would certainly present experimental difficulties but from a theoretical point of view, there might be a lot to be gained in this way. Now in order to deal with these problems one has to have solutions for inhomogeneous media; that is the scattering of electromagnetic waves from a large molecule which has a variable refractive index. And problems of this sort have not received very much attention from those engaged in electromagnetic scattering theory. Most of the effort has been devoted either to boundary value problems with perfectly reflecting boundaries in free space or else to problems in which various homogeneous media are separated by boundaries. And with a few exceptions, practically no effort has been devoted to inhomogeneous media; that is media which have a continuously variable refractive index.

Now a second important point at which the classical theory could be useful is to serve as a model for solutions of the Schrodinger or Dirac equations which are the pertinent equations in dealing with the scattering of electrons or nuclear particles. Here, almost any information or any techniques can be developed in connection with the solu-

tion of any wave equation, electromagnetic, acoustic, or whatever would be of great help in atomic or nuclear physics. The equations have very similar structures. The main difference is a potential energy appears which plays an exactly analogous role to the variable refractive index in the acoustical or optical case. One example that is of considerable current interest and one which occupied me and others at Stanford very greatly lately, is the scattering of very high energy electrons from nuclei. In this case, the nucleus is represented as a charge distribution which assumed to be spherically symmetric, but not necessarily uniform with a sharp boundary. Rather, instead of assuming that the charge density as a function of radius has a sharp edge it might have some sort of a rounded shape. This problem can be solved and has been solved in the case in which the charge distribution and therefore the potential or in the optical analog, the refractive index, has spherical symmetry. However, there is another closely related, and very interesting problem, of the case in which the charge does not have spherical symmetry and is distorted into an ellipsoidal shape. This actually occurs; in

many nuclei the nucleus is not at all spherical but is distorted and has a very large electric quadrupole moment. With that much distortion methods applicable to the spherical symmetry problem simply do not work. We did calculations by perturbation theory, but we did not do them as well as we would have liked to.

Another thing which will be coming up shortly is the question of granularity in the nucleus; there is reason to believe that there are local inhomogenieties which also will affect the scattering.

Now these are problems in which ideas or techniques from electromagneting scattering theory could be of great help. These are very difficult things and I don't wish to minimize the difficulties. It may well be then, that attention paid to electromagnetic scattering problems in the case of inhomogeneous media will pay off first of all in connection with the structure of large molecules, secondly by analogy with the more elementary problems of electron and nuclear scattering and, finally, it might even be that this may prove in the end to pay for itself in the better design of microwave lens systems.

APPENDIX

A-1 ABSTRACTS OF CONTRIBUTED PAPERS—SCATTERING, DIFFRACTION, AND GENERAL MATHEMATICAL PAPERS

1. Diffraction by an Infinite Grating of Cylinders in the Resonance Case—S. N. Karp and J. J. Radlow, New York University. We consider the problem of an electromagnetic plane wave incident on a grating of conducting cylinders of arbitrary shape. The electric vector is parallel to the generators of the cylinders. We examine the approach to the limiting case in which the wave number k, spacing d and angle of incidence ϕ are related by the condition that $k^2 = (2\pi n/d + k \cos \phi)^2$, for some integer n. No restrictions are imposed on the relative magnitudes of the cylinder dimensions and wavelength, but we require that both be small as compared to the spacing. The shape of the cylinder is immaterial. The final formulas are expressed explicitly in terms of second and third order determinants involving certain scattered amplitudes of an *isolated* cylinder congruent to an element of the grating. The amplitudes in question occur when the isolated cylinder is excited by the incident field and by a plane wave travelling parallel to the direction of the grating. The predicted variation of intensity with wavelength in each spectral order is essentially different from what would be expected for wavelengths bounded away from resonance.

2. Diffraction of Electromagnetic Waves Caused by Apertures in Absorbing Plane Screens—H. E. J. Neugebauer, Adalia Limited, Montreal, Canada. This paper is an approach to the problem of diffraction by an aperture in a plane infinite screen which is not a perfect conductor.

The total electromagnetic field is split up into an unperturbed field (closed aperture) and a perturbation field. The perturbation field can be split up in two fields each satisfying Maxwell's equations, one being symmetrical in the tangential components of the electric vector, the other one in the tangential components of the magnetic vector. The first one may be called the symmetrical, the other one the antisymmetrical perturbation field. Similar conditions to those postulated by Bethel[1] for the perfectly reflecting screen can be derived for the symmetrical, $\overrightarrow{E}_{\text{pert, s}}$, $\overrightarrow{H}_{\text{pert, s}}$ and the antisymmetrical $\overrightarrow{E}_{\text{pert, a}}$, $\overrightarrow{H}_{\text{pert, a}}$ field.

Up to this point the theory which can be generalized to include transparent screens, is rigorous providing screen thickness is negligible compared to free space wavelength.

Approximate solutions for the perturbation field can be obtained for large apertures if the unperturbed field is known and multiple scattering can be neglected.

Two very simple cases are the perfectly reflecting and the perfectly absorbing screen, the incident field being plane and linearly polarized.

[1] H. A. Bethel. *Phys. Rev.*, vol. 66, p. 163; 1944.

An approximate solution obtained for the perfectly reflecting screen from Bethe's conditions is:

$$H_x(F) = O$$
$$H_y(F) = \frac{k}{2\omega\mu_0\pi} \int \frac{\partial}{\partial Z_A} \left[\frac{EXP(-jkr)}{r} \right] ds$$
$$= -\frac{k}{2\omega\mu_0\pi} \frac{\partial}{\partial Z_F} \int \frac{EXP(-jkr)}{r} ds$$

where A denotes a point in the aperture and F the field point behind the screen.

This solution, $\overrightarrow{E}_{\text{pert, s}}$, $\overrightarrow{H}_{\text{pert, s}}$, is symmetrical. It is in agreement with experimental results.[2-4]

Interchanging of the electric and magnetic vectors yields a solution for the fictitious case of a screen which is a perfect magnetic conductor. This solution, $\overrightarrow{E}_{\text{pert, a}}$, $\overrightarrow{H}_{\text{pert, a}}$, is antisymmetrical.

The solution for the perfectly absorbing screen is the arithmetic mean of that for the perfect electric and perfect magnetic conductor.

Without going into details is may be added that, for a screen whose reflection coefficient for normal incidence is ρ^2, the solution can be obtained in a form

$$\frac{1 + \rho}{2} \overrightarrow{E}_{\text{pert, s}} + \frac{1 - \rho}{2} \overrightarrow{E}_{\text{pert, a}}.$$

The described method obviously does not yield the field near the screen. Yet, it suggests that also in other cases of absorbing scatterers a useful far field approximation might be obtained by proper superposition of the solutions for the two limiting cases of a perfect electric and perfect magnetic conductor.

[2] H. E. J. Neugebauer, The Eaton Electronics Research Laboratory Report No. b6 under Contract AF 19(122)—81 to the USAF, Cambridge Research Center.
[3] G. Bekefi, *J. Appl. Phys.* vol. 24, p. 1123; 1953.
[4] S. G. Buchsbaum, A. R. Milne, D. C. Hogg, G. Bekefi, G. A. Woonton, "Microwave diffraction by apertures of various shapes," *J. Appl. Phys.*, to be published.

3. Microwave Tandem Slit Diffraction—Leroy R. Alldredge, U. S. Naval Ordnance Laboratory. The diffraction of a plane electromagnetic wave by two identical slits in tandem is investigated for normal incidence with the polarization parallel to the edges of the slits. The slits are assumed to be infinitely long thus reducing the problem to one in only two dimensions.

Theory shows that the scattering cross section coefficient (σ) is proportional to the imaginary part of the far field forward scattering factor. The stationary form of σ is developed in terms of the incident field and unknown currents on the edges of the conductors forming the slits.

Experimentally, a parallel plate system described earlier by Row is used. Measurements were made for tandem separations up to 2λ and for slit widths from 0 to 1.4λ.

For slit widths greater than 0.3λ and for a tandem slit separation of zero, corresponding to a single slit, σ computed from the stationary form assuming zero edge currents is in good agreement with the exact theory of Morse and Rubenstein and as determined experimentally.

A calculation was carried out using the Kirchhoff-type approximation in the stationary form of σ for a tandem slit separation of 0.157λ. The results are in good agreement with the experimental values for slit widths greater than 0.5λ. The experimental results show an interesting resonance effect as the slit width changes.

4. Convergent Representations for the Radiation Fields from Slots in Large Circular Cylinders—L. L. Bailin and R. J. Spellmire, Hughes Aircraft Company. It was pointed out by Sensiper[1] and others,[2] that cylindrical wave representations in terms of harmonic components were very slowly convergent when applied to large cylinders. This difficulty seriously handicaps the design of cylindrical antennas in the millimeter wavelength region since it makes the determination of the radiation characteristics from slots on cylinders where ka is large a tedious numerical job ($k = \frac{2\pi}{\lambda}$ and a is radius). It is the purpose of this study to present approximate representations for cylindrical waves which are accurate enough and suitable for use in the design of slot arrays antennas where ka is in the range of 10 to 1,000 or more.

Following the techniques of van der Pol and Bremmer,[3] highly convergent representations have been obtained which are valid in different azimuthal sectors around the cylinder. Sensiper[1] investigated the second order approximations to the principle radiation characteristics for both the circumferential and axial delta-slots and the present study extends these results to third order. This is necessary to achieve suitable accuracy for values of ka between 10 and 100. The accuracy of each representation is determined by comparison with known values computed from the harmonic series. Thus, the quantitative behavior of the approximations can be described in the so-called optically illuminated region where the azimuthal angle is given by $/\phi/ < \pi/2$, in the shadow region, $\pi \geq \phi > \pi/2$, and in the

[1] S. Sensiper, W. G. Stearns, and T. T. Taylor, "A further study of the patterns of single slots on circular conducting cylinders," TRANS. IRE vol. AP-3, p. 240; August, 1952. S. Sensiper, "Cylindrical Radio Waves," Hughes Technical Memorandum No. 310.
[2] W. Franz and K. Depperman, "Theorie der beugung am zylinder unter beruchtsichtigung der kreichwelle," *Annalen der Physik*, vol. 10, p. 361; 1952. W. Franz, "On the Green's functions of the cylinder and the sphere," *Z. fur Natur.*, vol. 99, pp. 705-716; 1954. I. Imai, "The diffraction of electromagnetic waves by a circular cylinder," *Z. der Physik*, vol. 137, pp. 31-48; 1954.
[3] H. Bremmer, "Terrestrial Radio Waves," Elsevier, New York.

transition region where $/\phi/ \sim /\pi/2$. By combining the efforts of Sensiper and the present study, it is felt that a convergent representation could be found which would accurately approximate the intractable harmonic series for large ka.

Electrodynamics of Continua—P. C. Rosenbloom, University of Minnesota. Maxwell's equations are supplemented by one more vector equation, similar to the Navier-Stokes equations, obtained by equating the rate of change of the momentum in a region to the Lorentz force acting on that region. The system now becomes a determined system of nonlinear equations of hyperbolic type related to the theory of plasmas considered in recent years by some physicists. We obtain *a priori* estimates for the solutions and an existence proof, also a mathematical foundation for the application of perturbation methods. We apply these results to the theory of the physical constants of a dissipative medium and to the theory of the Hall effect.

An Analysis of Edge Behavior in Vector Diffraction Theory—S. N. Karp, New York University. The problem of diffraction of an electromagnetic field by a plane disc of arbitrary shape is discussed by a procedure based on the electric or magnetic field itself. Each Cartesian component can be expressed in terms of a suitable acoustical field plus a radiating singular homogeneous solution.

An expansion theorem relative to the edge behavior of the acoustic solution will be presented. This allows the unique determination of that solution from boundary conditions alone, if these are assumed to be adopted uniformly. A general construction for singular functions of the required type is given in terms of the scalar Green's or Neumann's function of the disc for an arbitrary distribution of the characteristics singularity along the rim. For the half plane one obtains a particularly simple result. Certain geometrical relations are deduced between the leading terms of the singular field components, in terms of the arclength variation of curvature along the rim. These relations generalize known results for a circular disc, and lead to a functional equation for the determination of the arclength distribution of singularity. From the solution of this equation (where possible) and from the Green's and Neumann's functions, the necessary singular corrections to scalar theory can be constructed by quadratures.

7. Experimental Measurement of Diffraction of Light at a Half-Plane—Franklin S. Harris, Jr. and Glen J. Morris, University of Utah. Diffraction of light and other electromagnetic waves at a half-plane is of interest because of its geometrical simplicity and importance. Though Fresnel worked out a simple theory which is adequate for most cases, rigorous and approximate solutions by Sommerfeld and successors have lacked good experimental measurements for comparison. Earlier work at the University of Utah[1] has been extended with an improved recording electron multiplier photometer.

[1] Keith Leon McDonald and Franklin S. Harris, Jr., *J. Opt. Soc. Amer.*, vol. 42, p. 321; 1952.

With a long light path a variety of experimental conditions, including different screen boundary conditions, were investigated and the intensity distribution in the diffraction pattern measured. The implications of the results are discussed with relation to other experimental work and theories.

8. An Expansion Theorem for Electromagnetic Fields—Calvin H. Wilcox, Air Force Cambridge Research Center. If $E(r)$, $H(r)$ is a free-space electromagnetic field which is defined in the exterior of a sphere $/r/ = c$ and satisfies Silver's radiation condition then E and H have expansions of the form

$$A(r) = \frac{e^{ikr}}{r} \sum_{n=0}^{\infty} \frac{A_r(\theta, \phi)}{r^n}$$

valid in $|r| > c$. The vectors $A_n = A'_n \hat{r} + A^2_{n\theta} \hat{\theta} + A^3_n \hat{\phi}$ are determined by the radiation pattern $A_o = A^2_o \hat{\theta} + A^3_o \hat{\phi}$ through the recursion relations

$$i\,kA_1' = \frac{-1}{\sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta\,A_o{}^2) + \frac{\partial A_o{}^3}{\partial\phi}\right],$$
$$2ikn\,A'_{n+1} = n(n-1)\,A'_n + DA'_n,$$
$$n = 1, 2, 3, \cdots$$

where $Df = \dfrac{1}{\sin\theta}\dfrac{\partial}{\partial\theta}\left(\sin\theta\,\dfrac{f\partial}{\partial\theta}\right) + \dfrac{1}{\sin^2\theta}\dfrac{\partial^2 f}{\partial\phi^2}$,

while $2\,i\,k\,n\,A_n{}^2 = n(n-1)\,A_{n-1}{}^2 + DA_{n-1}{}^2 + D_\theta\,A_{n-1}$, $n = 1, 2, 3, \cdots$
and $2\,i\,k\,n\,A_n{}^3 = n(n-1)\,A_{n-1}{}^3 + DA_{n-1}{}^3 + D_\phi\,A_{n-1}$, $n = 1, 2, 3, \cdots$

where $D_\theta A_n = 2\dfrac{\partial A'_n}{\partial\theta} - \dfrac{1}{\sin^2\theta}A_n{}^2 - \dfrac{2\cos\theta}{\sin^2\theta}\dfrac{\partial A_n{}^3}{\partial\phi}$

and $D_\phi A_n = \dfrac{2}{\sin\theta}\dfrac{\partial A_n'}{\partial\phi} + \dfrac{2\cos\theta}{\sin^2\theta}\dfrac{\partial A_n{}^2}{\partial\phi} - \dfrac{1}{\sin^2\theta}A_n{}^3$.

As an application of the theorem a characterization of the free-space tensor Green's function

$$\Gamma^\circ(r, r') = \left\{ I + \frac{1}{k^2}\nabla\nabla \right\}\frac{e^{ik|r-r'|}}{4\pi|r-r'|}$$

is given. The free-space Green's function

$$G(r, r') = \frac{e^{ik|r-r'|}}{4\pi|r-r'|}$$

is characterized by
1. $(\nabla^2 + k^2)\,G(r, r') = O$, $r \neq r'$,

2. $\lim_{r\to\infty} r\left(\dfrac{\partial G}{\partial r} - i\,k\,G\right) = O$,

uniformly,

3. $\lim_{G\to o}\displaystyle\int_{s(r'\epsilon)}\dfrac{\partial G(r, r')}{\partial r}dS = 1$,

4. $G(r, r') = g(|r - r'|)$.

An analogous characterization of $\Gamma^0(r, r')$ is difficult to obtain because Γ^0 has no simple symmetry property like 4; however, the expansion theorem yields the following characterization: $\Gamma^0(r, r')$ is the unique tensor defined by $A(r) = \Gamma^0(r, r')\cdot\hat{u}$ where $\hat{u}$ is an arbitrary constant unit vector and

1. $\nabla\times(\nabla\times A(r)) - k^2 A(r) = O\,r \neq r'$,
2. $\lim_{r\to\infty} r\{\hat{r}\times(\nabla\times A)\,ikA\} = O$,

uniformly,

3. $A(r) \sim \dfrac{e^{ikr}\sin\theta}{4\pi\,r}\,\theta$, $r\to\infty$.

Here (r, θ, ϕ) are spherical coordinates with origin at r' and polar axis along $\hat{u}$.

9. The Hyperbolicity of Maxwell's Equations—W. James, University of Minnesota. The hyperbolicity of Maxwell's equations is verified according to Garding's criteria for hyperbolic equations.

The equations for dissipation media are transferred to the wave equation form and the Cauchy problem for these equations is solved by the M. Reisz method. That is, given the values of $\vec{E}$ and $\vec{H}$ at $t = o$, the solutions are found for $t > o$.

The above solutions are reduced to the special case for vacuum. It is proved that these solutions actually satisfy both the initial conditions and Maxwell's equations for vacuum. Crude estimates are worked out for the solutions, their time derivatives and curls. Crude conditions are obtained for the conservation of energy to hold.

Several special cases are worked out explicity for the homogenous Maxwell's equations.

The same method is applied to Dirac's equations for a free electron, and for the propagation of elastic waves.

10. Diffraction of 3.2 CM Electromagnetic Waves by Dielectric Cylinders and Semi-Cylinders—A. B. McLay and M. K. Subbarao,[1] Hamilton College, Canada. Diffraction patterns in the close vicinity of long, 1 inch diameter dielectric cylinders of lucite, tenite and hard rubber have been observed in planes through and behind the rods and transverse to the incident radiation, which was polarized parallel to the long axis of the cylinders. The patterns of lucite and tenite are in fair agreement with those calculated by Froese and Wait.[2]

Patterns have been observed similarly for a $1\frac{1}{2}$ inch diameter lucite cylinder and for the same rod machined to a plano-semi-cylinder and placed with plane face towards or away from the source. Calculations for the cylinder have not yet been made, nor have the more difficult ones for the semi-cylinder. Marked trends in the patterns in the three cases are correlated in part with geometrical optics focusing. In the case of the semi-cylinder with plane face away from the source, double internal reflection seems to cause observed distortion of the field. This is most marked near the bearing, from center of the plane face, predicted by geometrical optics for a beam totally internally reflected by the plane face and re-reflected at angles slightly less than critical by the cylindrical surface.

[1] M. K. Subbarao is now at the Indian Naval Laboratory, Cochin, India.
[2] Froese and Wait, *Can. J. Phys.*, vol. 32, p. 775; 1954.

11. Tensor Scattering Maxtrix for the Electromagnetic Field—D. S. Saxon, University of California. The scattering of an arbitrary incoming wave by a lossless, but otherwise unrestricted, scattering object is described in terms of a tensor scattering matrix, unitary in all of its indices—continous and discrete. This matrix can be explicitly exhibited

in terms of the *plane wave* scattering amplitude for each of two mutually perpendicular directions of polarization of the incident wave. In other words, once the scattering of a plane wave by a given scatterer is known for both polarizations, then the tensor scattering matrix is also (and immediately) known. With the aid of the tensor scattering matrix, the reciprocity relations are easily established in very general form and the cross section theorem, including an interesting extension, follows without difficulty.

12. A Simplification of Electromagnetic Scattering Problems Involving a Sphere— Nelson A. Logan, Air Force Cambridge Research Center. It is the purpose of this paper to emphasize the simplification that results from direct treatment of the problem of finding the radial components of the electric and magnetic field. This method is closely related to the auxiliary potential approach but has the advantage of dealing directly with components of the field from which the remaining components are readily obtained. The chief advantage is that simple vector-constructions are used to determine the radial components of the primary electromagnetic field. The application of the boundary conditions then leads readily to the desired scattered field in a manner that is no more difficult than the problem of scalar-wave scattering by a sphere. Application is made to the problem of a dipole oriented at right angles to the polar axis of a sphere of arbitrary electrical properties. By employing a well-known expansion theorem for a five-dimensional wave function, the radial components are readily found. The ease with which this solution is obtained suggests that the introduction of orthogonal vector wave functions or auxiliary potentials represents an unnecessary complication of a relatively simple problem.

13. Diffraction by a Semi-Infinite Cone— Leopold B. Felsen, Polytechnic Institute of Brooklyn. Alternative representations are obtained for the fields radiated by arbitrarily placed scalar (acoustic) and vector (electromagnetic) point sources in the presence of a semi-infinite cone with simple boundary conditions. A radial transmission representation is useful for the field evalua-

tion when either the source or observation point is located near the cone tip, and for the determination of the plane wave backscattering on the axis of cones having large or small apex angle. An angular transmission formulation is highly convergent for the computation of plane wave scattering by the tip of a small-angle cone, with incidence on or off the axis and observation points lying in a region which excludes the rays reflected from the sides of the cone according to geometric optics. The radial transmission representation is in terms of a discrete spectrum of modes, while the angular transmission formulation comprises a continuous spectrum. Approximate closed form results derived from the rigorous expressions above are presented and compared with results derived in the literature by physical optics and other approximate formulations.

14. Some Variational Formulas for the Changes in Electromagnetic Scattering Cross Section and Dyadic Green's Functions Due to Boundary Perturbations— Carson Flammer, Stanford Research Institute. The first part of this paper presents the first order change in the scattering cross section for electromagnetic waves resulting from a small deformation of the boundary of the scattering object.

The second part of the paper is concerned with the first and second variations of the dyadic Green's functions upon deformation of their original domain of definition. In particular, the variations of the dyadic Green's functions for a sphere and of those recently derived by the author for a circular disk are considered with a view towards obtaining good approximations to the dyadic Green's functions for oblate spheroids of large and small eccentricities.

15. Electromagnetic Scattering by Spheroids as Power Series in the Ratio Diameter/Wavelength— A. F. Stevenson, Wayne University. A general method previously developed by the author for calculating the scattered field as a power series in k is applied to the case of a spheroid, the electromagnetic constants of which are arbitrary. Results are presented for the differential and the total cross sections for arbitrarily-oriented spheroids and also for the cases

where the orientation is random or determined by streaming. Krishnan's relation $V_H = H_V$ is proved for a randomly-oriented scattering body of arbitrary shape.

16. Variational Corrections to Geometric Optics in Scattering by a Conducting Cylinder— Ralph D. Kodis, Harvard University. A pair of variational principles is formulated for obstacle scattering which are analogous to those already studied in connection with aperture scattering.[1] The first of these contains geometric optics implicitly while in the second the geometric optics term is made explicit and is split off to leave a stationary form for the correction alone. With the physical optics current distribution as a trial function, the first terms of the corresponding asymptotic series in inverse powers of ka can be evaluated for the cylinder scattering cross section. The relative accuracy of these results is estimated by comparing them with the known exact solution at frequencies for which $1 < ka < 10$, and it is pointed out that further improvement will require a more accurate trial function.

[1] H. Levine and J. Schwinger, *Comm. in Pure and Appl. Math.*, vol. 3, p. 355; 1950.

17. On the Correction to the Total Geometric Optical Scattering Cross Sections of a Circular Cylinder and of a Sphere— S. I. Rubinow, Harvard University. A scalar plane wave incident on a conducting circular cylinder is considered. The total scattering cross section has a well known series expansion in terms of Bessel functions which converges slowly for large values of ka. In this case the series may be summed approximately by means of two simplifications; firstly, replacing the Bessel functions by the dominant terms of the appropriate Debye asymptotic expansions (different for different values of index n relative to argument ka), and secondly, replacing the summation sign by an integration over the index. The result indicates that the correction to the geometric optical cross section so obtained is of order $(ka)^{-2/3}$. The results for the Dirichlet and Neumann boundary conditions show the characteristic difference in sign of the correction terms. Similar considerations pertain to the sphere.

A-2 ABSTRACTS OF CONTRIBUTED PAPERS—MULTIPLE SCATTERING, SCATTERING FROM ROUGH SURFACES, AND TRANSMISSION AND REFLECTION PROBLEMS

1. Optical and Radio Twilight and Modes— Thomas J. Carroll and Rose M. Ring, Massachusetts Institute of Technology. The brightness of the molecularly scattered twilight from the clear sky can be calculated roughly from the mode theory used in recent years to understand propagation of radio waves well beyond the horizon through the earth's troposphere which acts as a dielectric layer to guide waves feebly but reliably into the deep shadow. The idealized air layer is taken to be either a homogeneous layer, or a layer with linear, bilinear, or curvilinear taper to vacuum at some large height. Both the absolute value and the attenuation rate in db/mi of important modes come out for

idealized models in satisfactory agreement with observations both for radio and light waves without use of arbitrary parameters. Because the index of refraction of tropospheric air is substantially independent of frequency, the frequency dependence is small for the radio or optical twilight calculation over the vast spectral range involved (factor more than 10^8). The slight but essential novelty of the treatment is the use of both of the independent solutions of the radial wave equation in constructing the allowed modal wave functions in the inhomogeneous air layer, and fitting these continuously to the airless earth type of waves in the homogeneous vacuum above where the

usual outgoing radiation condition may be applied. By contrast, all current hypotheses about scattering of radio waves from turbulence in the troposphere assume unphysical discontinuities of the index or its derivatives throughout the medium.

2. Transmission Characteristics of Parallel Wire Grids with Variable Tilt Angle— O. J. Snow, U. S. Naval Air Development Center. Small diameter parallel conducting wires were imbedded in thin, flat plastic sheets and located close to a receiving paraboloidal dish using plane wave incident energy at X band. Polarization was parallel to the wires, whose grating interval was

aried between a fifth wavelength and a hole wavelength for different panels. Re-ieved intensity was measured for varied rating tilt angles about an axis lying par-llel to the wires and near the center of each anel. Sharp and intense absorption dips ere observed for tilt angles at which the arasitic reradiation maxima lay in the end-re direction. The shapes and angular posi-ions of transmission vs tilt angle curves are iven to a moderately close approximation y a tentative theory which assumes that e input impedance of the grating is con-ant with varying tilt angle and that the arasitic currents drawn by all wires are qual. The apparent absorption energy is urther assumed to be proportional to the reas under plots of the first order reradia-ion lobes. A more precise theory, which in-udes the effect of varying input imped-nce, was required to predict sharp absorp-ion dips of smaller magnitude, also observed xperimentally.

. Light Scattering of Colloidal Spheres—

V. Heller, Wayne University. Manual and ectronic computations of scattering cross ections for colloidal spheres were carried ut for the parameters $\alpha = 0.2$ (0.2) 7.0 1.0) 15 and $m = 1.05$ (0.05) 1.30, on the asis of the Mie-theory, and in conjunction ith L. Langden, J. McCartey, W. J. angonis and A. F. Stevenson. From these unctions the specific turbidities, scattering ntensities for selected angles, dissym-netries, wavelength exponents and depolari-ations were calculated. Analytical expres-ions were developed in order to allow inter-olations of the specific turbidities for any alue of α and m within the range given. elationships of particular interest, on ccount of the comparative simplicity, were ound (a) between the maximum turibity at given particle diameter or wavelength on he one hand and m on the other; (b) be-ween the α value at which the turbidity naximum occurs at either constant wave-ength or at constant particle diameter, and n. From a combination of these relation-hips it became possible to determine both he diameter and the refractive index of a iven colloidal material by means of tur-idity measurements alone. The range of alidity of the various analytical expressions vas found to extend up to m values of at east 2.0 and a few of them appeared to be uitable for extrapolations to m values of nterest in the radar range. The various elationships were compared with those ob-ained from approximating theories on ight scattering with particular emphasis on he treatment by van de Hulst.

. Solution of the Helmholtz Equation with Random Boundary Values—

Jack Kotik, Technical Research Group, New York, N. Y. Following Kame-de-Feriet, who dealt with the equations $u_{xx} + u_{yy} = 0$, $u_{zz} - u_t = 0$, we consider solutions of the wo-dimensional Helmholtz equation, in a half-plane, with random boundary values. We replace the usual boundary values $f(x)$ by $f(x, \omega)$ where $x \epsilon R$ (R = real numbers) and $\omega \epsilon \Omega$, [Ω a measure space with $u(\Omega) = 1$]. The fundamental assumption is hat $f(x, \omega)$ is measurable on $Rx\Omega$ and that

$\overline{|f|} = \int_{\Omega} |f(x, \omega)| \, d\mu$ is a bounded function of x. Some results are:

1) For almost all ω $f(x, \omega)$ generates a solution $u(x, y, \omega)$ of the Helmholtz equation via $u = \int_R k(x-x', y)f(x', \omega)dx'$, where $y > 0$ and $k(x, y)$ is an appropriate Green's function. $u(x, O^+, \omega) = f(x, \omega)$ outside a set of measure zero which may depend on ω.

2) $\overline{u(x, y)} = \int_{\Omega} u(x, y, \omega)d\mu$ is generated by $\overline{f(x)} = \int_{\Omega} f(x, \omega)d\mu$; $u(x, o^+) = \overline{f(x)}$ for almost all x.

If we assume further that $\sigma_a{}^2(x) = \overline{|f(x)|^2}$ is a bounded function of x, that $\overline{f(x)} = O$, and define

$\sigma^2(x) = \overline{f(x)^2}$
$\gamma(x, \xi) = \overline{f(x) f(\xi)}$
$\Gamma(x, \xi, y, \eta) = \overline{u(x, y) u(\xi, \eta)}$
$\int_a^2 (x, y) = \overline{|u(x, y)|^2}$
$\int^2 (x, y) = \overline{u(x, y)^2}$

we find that

$\Gamma(x, y, \xi, \eta) = \int_R \int_R k(x - x', y)$
$k(\xi - \xi', \eta) \gamma(x', \xi') dx' d\xi'$
$\left| \int^2 (x, y) \right|^{1/2} \leq \int_a (x, y) \leq \int_R |k(x - x', y)|$
$\sigma_a(x') dx'$

If $f(x, \omega)$ is stationary of order 2, i.e.,

$\gamma(x, \xi) = r(x - \xi)$

then

$\Gamma(x, y, \xi, \eta) = \rho(x - \xi, y + \eta)$, where

$\rho(x, y) = \int_R k(x - x', y) r(x') dx'$

These results may be useful in treating scattering from rough surfaces.

5. Forward Scattering for Non-Absorbing Mie Particles—

Rudolf Penndorf, Air Force Cambridge Research Center. Extensive computations have been completed for the angular Mie scattering coefficient i_θ using the IBM 701 electronic data processing machine. They have been computed for $0 \leq \alpha \leq 30$ in steps of $\Delta\alpha = 0.1$ (α = size parameter $2\pi r/\lambda$), for 5 indices of refraction between $n = 1.33$ and 1.50. The angular steps are 5° between $0° \leq \theta \leq 170°$ and 1° between 170° and 180°, where θ is defined as the angle between the incoming and the scattered ray, thus $\theta = 180°$ is the forward scattered ray.

The results show the already known fact that the flux scattered in the forward direction as compared to the total scattered flux increases with the size parameter α. But important details have been found, for example, whereas i_θ in the forward direction generally increases with α, two types of oscillations are superimposed. A large one leading to secondary maxima and minima at about the same α values where $k(\alpha)$ attains maxima and minima, and a smaller one in the interval between these secondary maxima and minima.

Furthermore, percentages of the flux scattered in the forward direction, *i.e.*, 180° $\pm \theta$, have been computed for $n = 1.33$ and selected values of α to show numerically the importance of the forward scattered flux and its dependence on the size parameter.

6. Total Mie Scattering Coefficients for Real Refractive Indices—

Rudolf Penndorf, Air Force Cambridge Research Center. Extensive computations have been completed for the total Mie scattering coefficient K using the IBM 701 electronic data processing machine, which show in detail the nature of the function. K has been computed for $0 \leq \alpha \leq 30$ in steps of $\Delta\alpha = 0.1$, for indices of refraction $n = 1.33, 1.40, 1.44$, 1.486 and 1.50. In addition computations have been carried out for $n \approx 1$ and $n = \infty$.

The results show minor oscillations superimposed on the large oscillations. From a careful analysis of all available data for the range from $n = 0.8$ to $n = \infty$ simplified approximate relations have been deduced to compute smoothed K values of an accuracy of 5 to 10 per cent for a very wide range of size parameters. Thus K values for any index of refraction and for any size range can be computed with an accuracy sufficient for most practical applications of the Mie theory.

7. A Method for the Calculation of the Distribution of Energy Reflected from a Periodic Surface—

William C. Meecham, University of Michigan. The problem consists of finding the amplitudes of the plane waves reflected from a periodic surface due to a plane wave incident upon the surface from a specified direction; consideration is restricted to two dimensional problems. Following the method of Trefftz, one can attempt to fit the boundary condition approximately through the use of a finite number of functions which satisfy the differential equation (here, the wave equation). One ordinarily proceeds through the use of a variational method (frequently based on the Rayleigh quotient). In the present calculations the square of the error in the boundary condition is minimized (rather than the Rayleigh quotient); furthermore, the class of trial functions is restricted to include only those plane waves, both homogeneous and inhomogeneous, diffracted from the periodic surface. One is led to consider the, in general, nonorthogonal set of functions formed when the diffracted wave functions are evaluated at the surface. In order to minimize the error in the boundary condition it is sufficient to construct an orthogonal set of functions from this nonorthogonal set. The Fourier coefficients of the incident function with this set then minimize the error. One can calculate the estimates to the reflection coefficients using these coefficients.

The problem has been programmed for MIDAC (the University of Michigan digital computer) in the above form. Numerical results will be presented and compared with the results of experiment.

8. Multiple Scattering by Randomly Distributed Obstacles—Methods of Solution—

Chiao-Min Chu, Detroit Institute of Technology, and Stuart W. Churchill, University of Michigan. Methods of solution of problems of multiple scattering of electromagnetic radiation by randomly distributed obstacles are evaluated with respect to generality, accuracy, and suitability for numerical calculations. Exact solutions are found to be limited to infinite regions and are generally practical only for isotropic scatter-

ing. Approximate solutions of the transport equation are limited to simple regions and are very complicated for anisotropic scattering. Approximate representation of the scattering process with the classical diffusion equation permits solution for almost all conditions, but the solution is inaccurate near the boundaries of the region of scattering. Approximate representation of the intensity by a number of discrete components permits solution for many geometries. Several two-component representations have been developed which yield more accurate results, particularly for anisotropic scattering, than diffusion theory. A six-component model appears to be considerably more accurate for most geometries but requires machine computations.

9. Atmospheric Attenuation of Solar Millimeter Wave Radiation[1-5]—H. H. Theissing

[1] J. H. Van Vleck and V. F. Weisskopf, *Rev. Mod. Phys.*, vol. 17, p. 227; 1945.
[2] J. H. Van Vleck, *Phys Rev.*, vol. 71, p. 425; 1947.
[3] T. F. Rogers, Cambridge Research Center Report E5078; October, 1951.
[4] R. G. Newton and T. F. Rogers, Cambridge Research Center Report E5111; November, 1953.
[5] Atmospheric absorption graphs presented by T. F. Rogers at the Navy Millimeter Wave Conference in Washington, 1953 (unpublished).

and P. J. Caplan, Evans Signal Laboratory. With a thermal detector in the focus of a 60 inch searchlight parabola that tracked the sun two quantities were measured.

1) The spectral intensity distribution of solar millimeter wave radiation for a well defined comparatively dry weather condition was evaluated by a wire mesh filter method, showing a sharp cutoff near .9 mm, a peak near 1.1 mm and a slower decay toward increasing wavelengths. (Presented at the New York meeting of the Optical Society April 8, 1955).

2) The total signal was measured under various weather conditions, making possible an experimental determination of the water vapor absorption for the above spectrum, taking into account balloon recorded data at various altitudes and the solar elevation angle. The attenuation coefficient obtained, when corrected to 1 per cent water vapor partial pressure at ground level, was 2.6 db/km.

If the solar signal were monochromatic, it would be easy to compare this value with the theoretical value for that wavelength in Van Vleck's graph, or in the graph of T. F. Rogers, AFCRC, which was computed ac-

cording to the Van Vleck-Weisskopf theory. However, since the whole spectral range was involved, a laborious computation was required to make a comparison with theory. The final results show that in order to obtain agreement with the experimental findings it was necessary to raise the theoretical water vapor attenuation by a factor of 2.4 throughout the window region of the spectral range.

10. Approximate Calculations for Light Scattering when the Refractive Index is near Unity—A. F. Stevenson, Wayne University. A variational principle is given which determines the differential cross section for light scattering by an arbitrarily shaped body, and this is applied to the case $m \sim 1$. The variational principle is equivalent to one used by Cohen (with a different approach) except for a correction in the interpretation of the Green's function. It is shown that the total cross section is simply related to the differential cross section for forward scattering. The connection with the methods of Rayleigh-Gans and Debye is pointed out. Detailed calculations are made for scattering by a sphere (as a check) and by a circular cylinder of arbitrary dimensions.

A-3 ABSTRACTS OF CONTRIBUTED PAPERS—WAVEGUIDES, PROPAGATION, AND SLOW WAVES AND SURFACE WAVES

1. Waveguide Impulse Response[1]—George I. Cohn, Illinois Institute of Technology. This paper develops the solution for the transient propagation of electromagnetic fields in a perfectly conducting cylindrical waveguide, of arbitrary cross section, filled with a lossy medium, when an impulse is applied at the input. The impulse response is of particular interest because its relative simplicity, compared to the response from any other type of input function, allows the build up and propagation phenomena to be readily pictured from the derived equations without the need of elaborate numerical computations. The basic results obtained are not restricted to the case for an impulse function input, since the response to any input may be obtained by superimposing the responses of impulse function inputs.

The various electromagnetic field components propagate in different manners. Consequently, the conventional steady state mode structure does not exist during the transient. This paper will also illustrate how the steady state mode structure evolves out of the transient propagation. The propagating signal leaves an electromagnetic wake which constitutes distortion. Examination of this wake provides information about the concept of group velocity.

[1] The principal references for this subject are two Ph.D. theses, M. Cerrillo, "Transient Phenomena in Waveguides," M. I. T., 1948; and G. I. Cohn, "Electromagnetic Transients in Waveguides," I. I. T., 1951.

2. On the Eigenvalue Problem of Slow-Wave Propagation in Cylindrical Structures—F. E. Borgnis, Harvard University. It can easily be shown that the problem of propagation of electromagnetic waves the phase

velocity of which is less than the velocity of light, leads to scalar wave equations $(\nabla^2 + \Lambda_\mu) \phi_\mu = 0$ in two dimensions, where the eigenvalue Λ_μ is required to be negative. A proper boundary condition yielding negative eigenvalues is $\sigma\phi_\mu + \partial\phi_\mu/\partial n = 0$, where σ is negative, if n denotes the outward normal of the boundary. This point of view permits a number of general statements on the properties of slow-wave propagation and on the physical conditions under which such waves occur.

3. Propagation of Transient Fields from Dipoles near the Ground—H. Poritsky, General Electric Company. A study is made of the propagation in air and underground of the electromagnetic field orignated by a transient current $i(t)$ in a short antenna (dipole) of moment $F(t)$, located at the boundary $z = 0$ of a flat earth.

The treatment is based on a resolution of a spherical wave into proper plane waves proceeding in various directions. This is applied to the incident spherical wave from the dipole placed at $z = h > 0$; each plane wavelet is reflected and refracted at $z = 0$, and the resulting wavelets are superposed, then h is allowed to approach zero. There results a double integral representation for the field at a point P. Of the two integrations one turns out to be independent of F, and leads to the mean f of the reflection coefficient over a cone of directions of semi-vertical angle n with OP as axis; this integral is, in general, elliptic. The second integration does involve F; for the case of a current pulse, this integral is evaluated explicitly both in air and over a certain solid cone underground in terms of $\overline{f}(\cos p)$, $\cos p = ct/R$.

The above highly abbreviated and somewhat inaccurate summary omits all mention of the complex elements in the solution which occur both in the directions of the normals of the plane waves and in their amplitudes [even for real $i(t)$] and permeate the analysis throughout.

4. Theory of the Multipath Propagation of Frequency-Modulated Electromagnetic Waves—John P. Vinti, Aberdeen Proving Ground. A general treatment is given of the propagation and reception of multitone frequency-modulated waves received directly by an antenna and indirectly by reflection from an infinite plane reflector arbitrarily located and oriented. The transmitting and receiving antennas are both allowed to have arbitrary orientations and to have as much generality in structure as is consistent with the following assumption: with reception of the direct ray alone, the receiver output faithfully reproduces the input signal that is modulated on to the carrier.

The multipath problem then becomes essentially as easy as that for the multitone case of two parallel dipole antennas. The only difficulties of the problem as outlined that do not occur with the dipole antennas are confined to the calculations of the noise signal ratio R. The latter is the ratio of the received signal produced by the reflected ray to that produced by the direct ray, and it turns out to be a function of many parameters.

The properties of the detected signals are examined in detail for $R < 1$, $R > 1$, and $R = 1 - 0$, in which latter case the multipath phase errors can be expressed in closed form, in terms of solutions of a certain transcendental equation.

Mode Conversions in Multimode Waveguides

Mode Conversions in Multimode Waveguides—Wesley Ayres, Sylvania Electric Products Inc.[1], and E. T. Jaynes, Stanford University. Multimode operation of waveguide transmission systems would yield lower transmission losses and larger bandwidths than dominant mode operation. The problem of mode conversions in a multimode waveguide is investigated theoretically for arbitrary distortions of a waveguide wall, for uniformly tapered sections, and for a random distribution of wall distortions all in a perfectly conducting metallic waveguide. Several general results are obtained and a number of specific problems solved.

[1] Work performed at Stanford University.

Propagation in Ferrite-Filled Transversely Magnetized Waveguides

Propagation in Ferrite-Filled Transversely Magnetized Waveguides—P. H. Vartanian, Sylvania Electric Products Inc.[1], and E. T. Jaynes, Stanford Univeristy. A solution to the problem of propagation of higher modes in a transversely magnetized ferrite-filled rectangular waveguide has been found. The dependence of the fields in the direction of magnetization is the same as for the classical modes. The dependence in the other transverse direction is given by the solution of a simple constant-coefficient linear differential equation.

[1] Work performed at Stanford University.

An Extension of Maxwell's Solution of the Wave Equation for Concentric Strata to include Tilted and Wavy Strata

An Extension of Maxwell's Solution of the Wave Equation for Concentric Strata to include Tilted and Wavy Strata—Paul B. Taylor, Wright Air Development Center. Maxwell's solution of the wave equation in spherical coordinates with the index of refraction varying radially has been found practicable in explaining the anomalies of air-to-air radio propagation at microwave frequencies. It is now found that the solution can be extended to include cases in which the gradient of the index is not strictly radial, and, in particular, the case of a tilted or wavy stratum of abnormal gradient in an otherwise normal atmosphere. The results are successfully applied to treat the anomalies of air-to-ground propagation. The irregularities in the dirtribution of signal intensity are due to the existence in the atmosphere of narrow strata in which the variation of index with height is nonlinear.

Some Variational Principles for Resonators and Waveguides

Some Variational Principles for Resonators and Waveguides—A. D. Berk, Hughes Aircraft Company. Variational expressions are derived for the propagation constant of a waveguide and for the resonant frequency of a cavity directly in terms of the field vectors for situations ordinarily not covered in the literature. These situations occur, in general, when the problem is not susceptible of a scalar formulation and are characterized by the presence of an anisotropic and/or inhomogeneous material such as an oriented ferrite.

Three types of variational principles are obtained: a) those expressed in terms of the electric field vector alone, b) those expressed in terms of the magnetic field vector alone, c) those expressed in terms of both field vectors. Admissibility conditions for trial functions depend on the type of the variational principle used. All variational expressions for the resonant frequencies of a cavity are explicit. Of the expressions for the propagation constant, that of type c) is explicit, while types a) and b) are in the form of a quadratic equation.

Problems where the above variational principles were applied, include: determination of the cutoff frequencies of a rectangular waveguide with dielectric slabs; determination of the propagation constant in rectangular waveguides with ferrite or dielectric slabs and in circular waveguides with dielectric or ferrite rods; calculation of the shift in the resonant frequency of a cavity when the boundaries are perturbed. The results compare favorably with existing exact solutions and are obtained with considerably less effort. In some cases they are simple enough to be used directly as design equations.

9. Contribution to a General Transmission and Matching Theory for Waveguides

9. Contribution to a General Transmission and Matching Theory for Waveguides—F. J. Tischer. A general energy theory for waveguides and their interaction with connected elements is derived without use of the impedance concept of the low frequency transmission line theory. With the resulting relations, the behavior of waveguides, the interaction between waveguides of different properties and between waveguides and connected equivalents of two- and four-terminal networks are described.

The relations derived between the electric and magnetic energies stored in the waveguide per unit length, the power flow and the reflection coefficient permit not only the description of the phenomena in the waveguide but also can be used to calculate a general relation for the interaction between waveguide and a connected cavity element. When applied to a cavity element terminating a waveguide the reflection coefficient ρ is:

$$\rho = \pm \sqrt{1 - \frac{N'v}{N_o} - \left(\omega \frac{\Delta W_m'}{N_o}\right)^2} + \omega \frac{\Delta W_m'}{N_o},$$

wherein N_o is the power transported in the forward wave in the waveguide, N'_v the dissipated power and $\Delta W'_m$ the surplus of oscillating magnetic energy in the terminating element. Corresponding relations can be written for elements connected to two or more waveguides. Discontinuities in waveguides can be considered as such elements. The requirements for compensation of discontinuities can be computed in this way. Of particular interest are fundamental relations describing the frequency dependence of the complex reflection coefficient for a terminating element. The application of the energy concept permits the analytical solution of the problem of matching waveguides of different cross sectional structures. If the cross sectional differences between the connected waveguides are small, the relations can be simplified. These simplified relations permit the interpretation of the physical situation in the interaction between the waveguides. The relations can be used further to derive formulas for the transmission in nonhomogenous waveguides.

10. The Application of Higher Cavity Resonance Modes to the Measurement of Free Electron Densities and Diffusion Coefficients

10. The Application of Higher Cavity Resonance Modes to the Measurement of Free Electron Densities and Diffusion Coefficients—K. S. W. Champion, Tufts University. The method of measuring free electron densities (such as in an ionized gas) by means of the shift in resonance frequency of a microwave cavity was first introduced by Brown and Biondi[1]. However, their work and subsequent investigations by others has always used the fundamental TM or TE mode of the cavity. For certain applications, such as measuring diffusion coefficients in a uniform magnetic field, this limitation to the fundamental cavity mode is not satisfactory. Thus the possibility of using higher resonance modes was investigated. For some measurements in progress, with along cylindrical cavity, the TE_{119} mode was chosen. This mode was chosen both because the electric field had a suitable configuration and because the resonant frequency lay in a range that could be readily measured.

The change in resonant wavelength is related to the electron density by the expression

$$\frac{\Delta\lambda}{\lambda^3} = - \left[\frac{e^2}{8\pi^2 m E_o C^2}\right] \frac{\int_{tube} nE^2\, dV}{\int_{cavity} E^2\, dV}$$

where n is the electron density per cm^3 and the other symbols have their usual connotation. The solution of the integrals where the electric field E has the value appropriate to the TE_{119} mode will be outlined. Test values of diffusion coefficients (without a magnetic field) measured by this method agree with values obtained by other methods.

[1] M. A. Biondi and S. C. Brown, *Phys. Rev.*, vol. 75, p. 1700; 1949.

A-4 ABSTRACTS OF CONTRIBUTED PAPERS—FERRITES, PLASMA OSCILLATIONS, AND ANISOTROPIC MEDIA

Field Displacement Isolators at 55 KMC

Field Displacement Isolators at 55 KMC—E. H. Turner, Bell Telephone Labs. Resistance sheet field displacement isolators[1] have been investigated analytically

[1] A. G. Fox, S. E. Miller, and M. T. Weiss, *Bell Syst. Tech. Jour.*, vol. 34, p. 5; January, 1955.

and experimentally. We have demonstrated that such isolators can be built at 55 kmc with reverse-to-forward loss ratios of greater than 100 to 1 in db over a range of greater than 2 kmc. The surprisingly good behavior of these isolators is due to the existence of an electric field null at the position of the resistance sheet. The necessary conditions on the tensor permeability, and hence on the magnetic properties of the ferrite, for this null to exist, will be demonstrated.

The electric field null condition has been

examined analytically and experimentally for various positions of the ferrite in the waveguide, for different ferrite dimensions and for various frequencies. The analysis of a more general situation where there is no electric field null at the ferrite has been carried through in particular cases. With the help of these solutions one can give a qualitative explanation of the behavior of the isolator in the high loss direction of propagation.

2. Use of Perturbation Theory for Cavities and Waveguides Containing Ferrites—G. S. Heller and Benjamin Lax, Massachusetts Institute of Technology.

Boundary value problems for electromagnetic propagation in ferrite media are exceedingly complicated and solutions from which useful information can be extracted exist for only a few isolated cases.

One can obtain perturbation formulas using exact expressions which yield the difference in complex eigenvalues for two different field configurations in a cavity or waveguide. From these, one can compute changes in Q and resonant frequency for a cavity or changes in attenuation and phase for a waveguide due to the insertion of ferrites which perturb the fields only slightly. Cases in which two or more degenerate modes are coupled by the ferrite can be handled by introducing suitable linear combinations of these degenerate modes of the empty cavity or waveguide.[1,2] Completely filled guides or resonators can be treated in frequency ranges for which the off-diagonal components are small compared to the diagonal components of the permeability tensor. Better convergence can be obtained if solutions for a medium with a diagonal permeability tensor are used as the unperturbed solutions.

Applications to problems not yet treated theoretically in the literature will be discussed and the results extended to magneto-ionic media in which the ionic density distribution is nonuniform.

[1] B. Lax and A. D. Berk, 1953 IRE Convention Record, Part 10, March 10, 1953.
[2] A. D. Berk, Q. P. R., Res. Lab. Elec., M. I. T.; January, 1954.

3. Propagation and Magnetoplasma Effects in Semiconductors—Benjamin Lax and Laura M. Roth[1], Massachusetts Institute of Technology.

The extension of the classical magneto-ionic theory to semiconductors has very interesting ramifications. The results differ from those for a simple plasma in that the magnetoplasma of a semiconductor usually possesses directional properties. These are a consequence of the band structure of such crystals as germanium and silicon where the energy surfaces in the Brillouin zone can be approximated by ellipsoids or degenerate fluted surfaces possessing cubic symmetry. The conductivity tensors, evaluated in a coordinate system in which the magnetic field is along one axis, involve certain parameters of the energy surfaces. The resulting conductivity tensor is substituted into Maxwell's equations and the propagation constant is evaluated for the magnetic field along three principal

[1] Now at Harvard University, Cambridge, Mass

crystal directions. Corresponding equations for the plasma oscillation frequencies are also derived. A simplified result for holes in germanium and silicon is considered when the directional effects are neglected. Finally a brief discussion will be made of depolarizing effects of the plasma encountered when finite geometries are involved.

4. Interaction between Plasma Oscillations and Electromagnetic Waves. I. Coupling Conditions[1]—R. M. Gallet, National Bureau of Standards.

An important part of cosmic radio emissions seems capable of explanation by plasma oscillations excited in shock-wave fronts. However, an understanding of the transfer mechanism of this energy into radiated electromagnetic waves must be sought, and the coupling conditions in the presence of a static magnetic field are examined. Coupling is found to be effective for the wave for the case of extraordinary transverse propagation, but only if the magnetic field is greater than a minimum value. This imposes the interesting condition that the magnetic energy density must be greater than the total kinetic energy density of the plasma charges. Various astrophysical situations are examined in the light of this result. The precise frequency at which coupling takes place and a very simple expression for the bandwidth are also obtained. It is shown that the electromagnetic wave is enclosed inside the shock front by internal reflection, so that the coupling permits the build up of a great electromagnetic energy density. Very low partial transmission is expected at the border.

[1] It is intended that this work shall be continued and a second part is planned as follows: II. Study of the Transmission Factor.

5. Influence of the Irregularities of Land on the Propagation of Radio Waves—Especially at Great Distances Beyond the Horizon—J. Voge, Laboratoire National de Radioelectricite, Paris, France.

The object of the author is to calculate the field strength received at a great distance beyond the horizon, in homogeneous or standard atmosphere, above an irregular surface. The method, inspired by AMENT (San Diego Symposium, 1953), consists in making a conformal transformation which reduces the problem to the case of a propagation along a regular surface, but in a slightly heterogeneous atmosphere, the characteristics of the ground being themselves slightly heterogeneous. The case of irregularities extending along a given direction (one-dimensional irregularities) has been especially considered, but the results have been generalized for a fully random surface (two-dimensional irregularities).

Calculation of the field at a great distance is thus made equivalent to the one of a field scattered by a portion of heterogeneous atmosphere, which is to be added to the field due to normal diffraction over an even surface. It is found that this "scattered" field varies, as a function of the wavelength λ, as $\lambda^{0.5}$. The field diminution with the distance D comprises a decrease proportional to $D^{-4.5}$ and additional variations depending on the exact profile of the irregularites or undulations of the surface. Different special

cases have been examined (sinusoidal undulations, single hump, surface with irregularities variable in time such as the surface of the sea).

6. Microwave Single-Sideband Modulator using Ferrites—J. C. Cacheris and H. A. Dropkin, Diamond Ordnance Fuze Laboratories.

This paper describes a single-sideband modulator for shifting the frequency of an X band signal by means of a rotating magnetic field transverse to a ferrite differential half-wave section. The device is one of the first practical applications of the double-refraction properties of ferrites.

Improvements over an earlier model include reduction in size and continuous operation without drift. An efficient and compact magnetic structure has been designed for producing the rotating magnetic field. Excessive heating of the ferrite and voltage breakdown of the coils is eliminated by a forced-air cooling system.

The modulator shifts the microwave carrier frequency of 9,375 mc by plus or minus 20 kc. With a rotating field of approximately 200 oersteds the microwave insertion loss is 1.0 db. The undesired sideband suppression is above 40 db while the carrier suppression is 23 db. For a frequency bandwidth of 500 mc, the insertion loss remains below 40 db.

7. An Electric Dipole above an Infinite Anisotropic Slab—Bernard Friedman, New York University.

This paper investigates the field produced by a point electric dipole situated in air above an infinite horizontal anisotropic uniaxial slab of thickness d. We assume that the two equal dielectric constants of the slab are the same as that of air. It is found that, besides the usual space-type field decaying as the inverse power of the horizontal distance, there exists a surface wave which decays as the inverse square root of the horizontal distance except in a certain dihedral angle around the direction of non-isotropy of the slab. In this dihedral angle the field is only a space wave and has no surface wave component. The size of this dihedral angle varies inversely with the thickness of the slab.

8. Nonlinearity of Microwave Ferrite Media—N. G. Sakiotis, H. N. Chait, and M. L. Kales, Naval Research Laboratory.

Existing theories of propagation in magnetized ferrite media predict propagation constants which are independent of the rf field strength only if a number of restrictive conditions are satisfied. In general, however, the propagation constants are functions of the rf field strength and nonlinear propagation is to be expected. One of the conditions for linearity is that the rf magnetic field intensity be small compared to the static magnetizing field intensity. This condition can be violated when the peak power level of the wave incident on the ferrite medium is sufficiently high and indeed, it has been observed by the authors that ferrite loaded waveguides can become nonlinear at peak power levels as low as 1 kw.

Results are presented of a study of the behavior of ferrite loaded waveguides at 9,400 mc over input peak power levels from .1 kw to 100 kw which includes power level

ommonly encountered in radar applications. The dependence upon input power level of the ferrite losses, phase shift, and

rotation of the plane of polarization is described for a number of ferrite materials. The effect upon the degree of nonlinearity

of the intensity of the static magnetizing field is discussed as well as the dependence upon the dimensions of the ferrite.

A-5 ABSTRACTS OF CONTRIBUTED PAPERS—ANTENNAS AND MICROWAVE OPTICS

Phase Centers of Microwave Antennas— David Carter, Convair, General Dynamics Corporation. This paper is concerned with the location of the phase centers of microwave antennas. The inadequacy of conventional aperture theory for the accurate description of phase centers is discussed. Formulas are developed and, for numerical indications, calculations are made for paraboloidal reflectors of different f/D ratios and a class of primary patterns which provide an approximate representation of a great many common feeds. The results are presented in graphical form to provide useful design information and show the dependence of principal E- and H- plane phase center location on feed and dish parameters. Contrary to the prediction of aperture theory, it is shown that the phase centers of axially symmetric antennas are not in the aperture plane, but they are dispersed about it.

A Method of Analyszing Coupled Antennas of Unequal Sizes— C. A. Levis and L. T. Tai, Ohio State University. The simultaneous integral equations involved in the formulation of coupled antennas of unequal sizes can be solved by a variational method. The unknown coefficients pertaining to the trial functions for the current distributions on various elements are shown to satisfy a set of linear equations. The method is directly applicable to analyze Yagi-Uda Antennas, and similar types of linear antenna systems. The same technique can be applied to coupled slots inside a waveguide or radiating slots excited by waveguides.

Diffraction of Surface Waves by a Semiinfinite Dielectric Slab— Carlos M. Angulo, Brown University. The discontinuity at the end of the slab is regarded as the junction of an open dielectric-filled waveguide and a free-space waveguide.

A variational expression is set up for the terminal impedance representing the effect of the discontinuity on the surface wave. Close upper and lower bounds for the impedance are obtained.

The transfer impedance between the surface wave and, either a) one of the modes of the continuous mode spectrum associated with the free-space wave-guide, or b) one of the modes of the continuous mode spectrum associated with the dielectric guide, is expressed in terms of the terminal impedance of the slab. The transfer impedances yield the modal amplitudes of the fields in each waveguide.

The synthesis of all the modal components gives the electromagnetic field scattered forward and backward by the end of the slab. The synthesis is carried out for the far field by the method of steepest descents.

Terminal impedance and radiation patterns are plotted, as functions of the thickness, for a polystyrene slab ($\epsilon = 2.49$).

4. Radiation Conductance of Slots in Plane and Curved Conducting Surfaces—[1-3]James R. Wait, Defence Research Board, Canada and James Y. Wong, National Research Council, Canada. A comprehensive study is made of the radiative properties of narrow axial slots located on plane and curved metallic conducting surfaces. Expressions are developed for the far-zone fields of a narrow half-wave slot on a knife edge, right-angle wedge, thin ribbon, elliptic and circular cylinder. The Poynting vector method is employed to determine the radiation conductance, and extensive numerical results of slot conductance are presented.

It is shown that for the case of a plane surface, the conductance is an oscillating function of the distance of the slot from the edge and having a period of approximately one-half wavelength. For distances greater than 1.25λ, it is found that the conductance is within 3 per cent of the conductance of the same slot in an infinite flat ground plane. For the case of a circular cylinder, the conductance curve is a monotonically increasing function of radius. Numerical results are included for the surface current distribution on a thin ribbon in an endeavor to explain the behavior of the conductance curves. Some experimental work is provided to corroborate the theoretical results.

J. Y. Wong, "Radiation conductance of axial and transverse slots in cylinders of elliptical cross section," vol. 41, p. 1172; September. 1953.
[2] J. R. Wait and S. Kahana, *Can. Jour. Phys.*, vol. 32, p. 714; 1954.
[3] J. R. Wait and R. Walpole, *Can. Jour. Tech.*, vol. 33, p. 97; 1955.

5. Further Investigations of Aberrations in Microwave Lenses— M. P. Bachynski and G. Bekefi, McGill University, Canada. The effect of the presence of a mixture of aberrations upon the electromagnetic field intensity distribution in the image plane of rotationally symmetrical solid dielectric microwave lenses has been investigated experimentally at a wavelength of 1.25 cm. The results obtained are in agreement with the diffraction theory of optical aberrations which has been extended to include a number of aberrations present simultaneously. The field intensity distribution along the principal ray of the system is also described.

Some measurements of relative phase in the image planes of the microwave lenses have been made. These, together with associated problems are discussed.

6. Radiation by a Disk and Conical Structures— A. Leitner and C. P. Wells, Michigan State University. The scalar problem of radiation by a vibrating rigid disk (Kolbenmembran) has been solved by Bouwkamp[1] and Spence[2], using the spheroidal coordi

[1] C. J. Bouwkamp, Diss. Groningen; 1941.
[2] R. D. Spence, *J. Acous. Soc. Am.*, vol. 20, p. 380; 1948.

nates; by Sommerfeld[3] as a two-part boundary value problem, using cylindrical and spherical coordinates, in terms of an infinite system of linear equations. It is here solved, in spherical coordinates, by formulating it, with the help of the Lebedev[4] transform, as a set of dual integral equations on the complex order plane.

The method of solving these integral equations is an extension of Karp's solution of the electrostatic problems of disks and cones[5]. It involves the properties of Bessel functions as functions of order. In the procedure we use the device of Oberhettinger[6] of converting to imaginary propagation constants.

We are now investigating a certain class of two-part problems in antenna theory amenable to solution by this method; in particular, the radiation from finite hollow conducting cones and from un-capped bi-conical structures of arbitrary apex angle.

[3] A. Sommerfeld, *Ann. Phys.*, vol. 42, p. 413; 1942-1943.
[4] N. N. Lebedev, *Dokl. Akad. Nauk. SSSR*, (N. S.) vol. 58, p. 1007; 1947.
[5] S. N. Karp, N. Y. Univ. Math. Res. Group, Reports EM-25, vol. 35; 1950-1951.
[6] F. Oberhettinger, *Comm. Pure & Appl. Math.*, vol. 7, p. 551; 1954.

7. The Fresnel Field of a Finite Line Current Distribution— R. B. Barrar and C. H. Wilcox, Hughes Aircraft Company. Increased resolution requirements on microwave antennas have led to the use of long linear arrays at very short wavelengths. The possibility exists that such antennas will be tested, and even used, in their Fresnel field regions.

Using a method first mentioned by Sommerfeld[1] to expand the eletromagnetic fields in powers of l/r, the fields and power patterns of any reasonable current distribution can be calculated rather simply in regions of r, the distance from the line source, well within the classical far field region. Numerical calculations of phase amplitude and power patterns have been made for the following sets of parameters:

a) Current distribution: cosine-on-a-pedestal, with 10, 15, and 20 db power taper.

b) $u = \dfrac{kL}{2} \cos\theta = 0(0.2)10$, where θ is the azimuth angle from broadside, and L is the length of the line source.

c) $r = qL^2/\lambda$, where $q = \infty$, 2.0, 1.0, 0.5, and 0.25.

It was found that the first five terms of the power series are sufficient to obtain accurate phase and amplitude results as far in as $r = L^2/2\lambda$, as compared with the classical far field transition region of $2L^2\lambda$, but

[1] A. Sommerfeld, "Partial Differential Equations in Physics," Academic Press, New York, 1949, p. 19.2

that further terms become necessary as r decreases. Empirical expressions are obtained for beam broadening, loss of gain, sidelobe level deterioration, and other Fresnel field effects. The Fresnel field is shown not to be independent of the far-field scaling parameter kL. This dependence, however, is negligible for $kL > 500$, (the region of principal interest in the present calculations) which fact permitted several simplifying approximations to be used.

8. Variable-Index Lenses Producing Conical Wavefronts—K. S. Kelleher, Melpar, Inc. There exist three spherical, variable-index-of-refraction lenses of some interest in microwave optics. In order of their appearance in the field, they are: the Maxwell (Fish-Eye), the Luneberg, and the Eaton lenses. The first images a point source into another point source; *i.e.*, spherical wave; the second images the point source into a plane wave; the third images the source into a cylindrical wave. The present paper shows that all of these lenses are members of a family of spherical lenses which produce conical wavefronts; these are wavefronts which coincide with sections of conical surfaces. For the Eaton lens, the cone angle is zero so that the cone becomes a cylinder; for the Luneberg lens, the angle is 180 degrees so that the cone becomes a plane; for Maxwell lens, angle becomes 360 degrees and the cone degenerates into a point.

INDEX TO AUTHORS

INSTITUTIONAL LISTINGS

The IRE Professional Group on Antennas and Propagation is grateful for the assistance given by the firms listed below, and invites application for Institutional Listing from other firms interested in the field of Antennas and Propagation.

COLLINS RADIO COMPANY, Cedar Rapids, Iowa
Antenna Design and Propagation Research Related for Airborne and Ground Communication Systems.

DEVELOPMENTAL ENGINEERING CORP., Washington, D. C. and Leesburg, Va.
Antenna Systems Research, Design, and Evaluation

DORNE AND MARGOLIN, INC., 30 Sylvester Street, Westbury, L. I., New York
Antenna Research and Development—Radiation Pattern Measuring Services.

D. S. KENNEDY & CO., Cohasset, Mass.
Microwave Antennas, Reflectors, Lenses, Radomes and Accessories. Design, Development and Production.

THE GABRIEL LABORATORIES, Div. of the Gabriel Co., 135 Crescent Road, Needham Heights 94, Mas
Research and Development of Antenna Equipment for Government and Industry.

HUGHES AIRCRAFT COMPANY, Culver City, California
Research, Development, Mfr.: Radar, Missiles, Antennas, Radomes, Tubes, Solid State Physics, Computers.

I-T-E CIRCUIT BREAKER CO., Special Products Div., 601 E. Erie Ave., Philadelphia 34, Pa.
Design, Development and Manufacture of Antennas, and Related Equipment.

JANSKY & BAILEY, INC., 1339 Wisconsin Ave. N.W., Washington 7, D.C.
Radio & Electronic Engineering; Antenna Research & Propagation Measurements; Systems Design & Evaluatio

MARYLAND ELECTRONIC MANUFACTURING CORPORATION, College Park, Md.
Antenna and System Development and Production for Civil and Military Requirements.

RADIO ENGINEERING LABS., INC., 36-40 37th St., Long Island City 1, N. Y.
Equipment for Communication and Propagation Test Beyond the Horizon UHF Systems.

WEINSCHEL ENGINEERING CO., INC., Kensington, Md.
Attenuation Standards, Coaxial Attenuators and Insertion Loss Test Sets.

WHEELER LABORATORIES, INC., 122 Cutter Mill Road, Great Neck, New York
Consulting Services, Research and Development, Microwave Antennas and Waveguide Components.

The charge for an Institutional Listing is $25.00 per issue or $75.00 for four consecutive issues. Application may be made to the Technical Secretary, The Institute of Radio Engineers, 1 East 79th Street, New York 21, N.Y.